

SynFog: A Photo-realistic Synthetic Fog Dataset based on End-to-end Imaging Simulation for Advancing Real-World Defogging in Autonomous Driving

Yiming Xie^{1*}, Henglu Wei^{1*}, Zhenyi Liu², Xiaoyu Wang¹, Xiangyang Ji¹

¹Tsinghua University, ²Stanford University

{xieym22@mails., weihenglu@mail., xiaoyu-w17@mails., xyji@}tsinghua.edu.cn,
 {zhenyiliu@}stanford.edu



Figure 1. The SynFog dataset comprises images under different lighting, with the first row captured in natural sky light and the second with extra light sources like street lamps and automotive lighting. This dataset covers three levels of fog density across each scene, corresponding visibility of 600 m, 300 m, 150 m. Additionally, pixel-accurate depth data and segmentation labels for each scene are also provided.

Abstract

To advance research in learning-based defogging algorithms, various synthetic fog datasets have been developed. However, existing datasets created using the Atmospheric Scattering Model (ASM) or real-time rendering engines often struggle to produce photo-realistic foggy images that accurately mimic the actual imaging process. This limitation hinders the effective generalization of models from synthetic to real data. In this paper, we introduce an end-to-end simulation pipeline designed to generate photo-realistic foggy images. This pipeline comprehensively considers the entire physically-based foggy scene imaging process, closely aligning with real-world image capture methods. Based on this pipeline, we present a new synthetic fog dataset named SynFog, which features both sky light and active lighting conditions, as well as three levels of fog density. Experimental results demonstrate that models trained on SynFog exhibit superior performance in visual perception and detection accuracy compared to others when applied to real-world foggy images.

1. Introduction

Fog, due to its light scattering and attenuation properties, poses a significant challenge for outdoor vision-based sys-

tems such as remote sensing, surveillance, autonomous driving, and fire rescue [6, 10, 52]. It causes severe degradation in object appearance and contrast, resulting in a notable impact on people’s daily lives. Consequently, there has been a growing research interest in image defogging task to develop robust outdoor vision systems.

Recent learning-based defogging algorithms depend on sets of foggy images and corresponding clear images for model training. While several datasets have been collected from real-world scenarios [5, 40], they are typically limited to foggy images only, as it is almost impossible to obtain pixel-accurate clear images with consistent scene contents and environmental illumination. Studies in [2–4, 35] attempt to capture real-world foggy and clear image pairs using smoke machines. However, the scale of these datasets is inadequate for training of a high-performing model. Given the challenge of obtaining paired foggy and clear images in real world, various synthetic fog datasets have been proposed as an alternative [25, 40, 55]. The most widely used method is based on the theory proposed by McCartney [33], also known as the Atmospheric Scattering Model (ASM). However, this model fails to consider accurate global illumination and the actual imaging process that occurs in real-world captures as detailed in Sec. 3.1. These limitations significantly impact the model’s overall applicability, par-

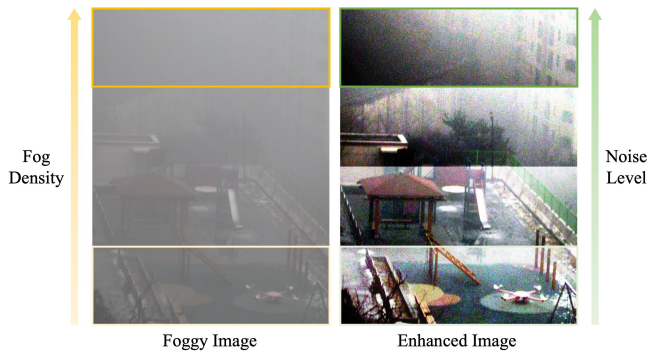


Figure 2. Relationship between noise level and fog density in real-world foggy images. Using regional contrast stretching for image enhancement, it can be observed that the noise level increases with the concentration of fog.

ticularly in nighttime scenarios, and can lead to a disparity between synthetic and real-world foggy images. As a result, models trained on synthetic foggy images may exhibit limited robustness when applied to real-world ones.

To address these issues, we propose an end-to-end foggy image simulation pipeline that encompasses the entire imaging process starting from 3D virtual scenes to final photorealistic foggy images. Our approach involves foggy scene rendering using volumetric path tracing and imaging through a realistic camera model. As shown in Fig. 3b, fog is treated as a spatial volume, and the spectral scene radiance is rendered using volumetric path tracing [38]. By defining the phase function of fog medium and bidirectional reflectance distribution function (BRDF) of the scene components, we are able to accurately describe the scattering process with multiple light sources under foggy conditions. The foggy scene radiance data is further processed through a physically-based camera model, which comprises optics, sensor and image processing, to faithfully replicate real camera devices. Based on the proposed pipeline, we develop a new extensive synthetic fog dataset called SynFog. This dataset contains 500 unique outdoor scenes, each featuring two kinds of lighting conditions and three levels of fog density. Extensive experiments have been conducted to qualitatively and quantitatively validate the authenticity of our simulation pipeline and the SynFog dataset.

Our main contributions are summarized as follows:

- We propose a physically-based end-to-end foggy image simulation pipeline that incorporates accurate light transportation in scattering medium and physical characteristics of optics and sensor into the synthesized images.
- We develop a photo-realistic synthetic fog dataset, SynFog, which includes both sky light and active lighting conditions, as well as three levels of fog density.
- We demonstrate the authenticity and effectiveness of our simulation pipeline and the superior realism of SynFog dataset compared to other synthetic fog datasets.

2. Related Work

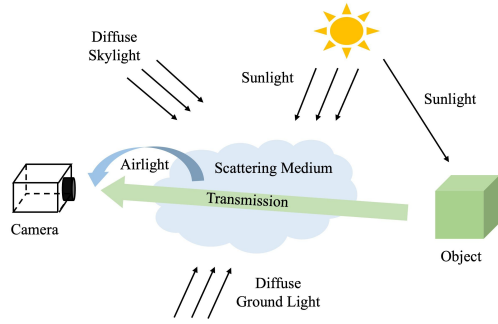
Given the challenge of acquiring paired foggy and clear images in the real world for training learning-based defogging methods, researchers have proposed many synthetic fog datasets as an alternative. In this section, we will offer a thorough review and analysis of foggy image simulation techniques and their associated datasets.

Based on Atmospheric Scattering Model (ASM). This method involves rendering fog onto clean images along with their depth information [1, 19, 25, 26, 40, 53, 55]. Depth information can be acquired through game engines, depth cameras or depth estimation methods. In an earlier study conducted by Tarel *et al.* [46], FRIDA was constructed using a total of 90 synthetic images, with clear images and depth maps generated using *SiVICTM* software. Subsequently, in their later work [47], more synthetic images were added, leading to an upgraded version named FRIDA2, which includes 330 synthetic images. However, the scale of these datasets is insufficient for modern data-driven defogging methods, and the foggy images appear unrealistic. In more recent works [26, 53], a similar approach has been employed to acquire both the clear images and their corresponding depth maps from game engines. In [25], a general dehazing dataset called RESIDE was introduced. The indoor training set (ITS) of RESIDE was constructed based on existing indoor RGBD datasets NYUv2 [42] and Middlebury stereo [41]. For the outdoor training set (OTS), 2071 real-world outdoor images were collected, and the corresponding depth information was estimated using a learning-based technique [27]. It is important to note that the accuracy of depth estimation based on learning methods is still unsatisfactory, as single-image-based depth estimation is inherently an ill-posed problem. Furthermore, Foggy Cityscapes was introduced in [40], utilizing images from the Cityscapes dataset [9]. Despite efforts to enhance the original depth maps in Cityscapes, artifacts and discrepancies still persist, resulting in synthetic foggy images that appear unnatural.

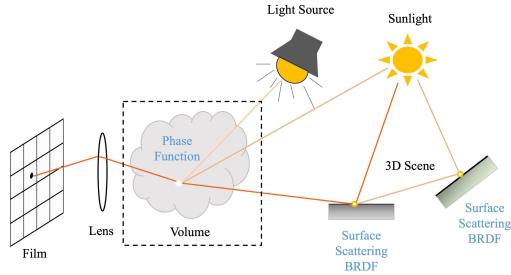
Based on Real-time Rendering Engines. Foggy images generated utilizing real-time rendering engines like Unity and Unreal have been discussed in works such as [18, 44]. Among them, Virtual KITTI 2 [18] stands out as a notable example. However, these foggy images often lack realism and exhibit limited diversity in terms of fog density.

Based on Physical Means. Fog/haze machines are employed in real-world settings to generate artificial fog, as discussed in works such as [2–4, 35]. However, these datasets are generally limited in scale and may not offer sufficient supervision for learning-based defogging methods. Moreover, ensuring consistent content between the clear and foggy image captures can be challenging.

There are also several real-world fog datasets available, such as Foggy Driving [40], Foggy Zurich [11], and Seeing



(a) Atmospheric Scattering Model (ASM) [34].



(b) Foggy scene rendering based on volumetric path tracing.

Figure 3. Foggy scene simulation methods.

Through Fog [5]. However, due to the absence of corresponding clear reference images, these datasets can only be used for testing purposes rather than for training. In this paper, we propose a physically-based end-to-end foggy image simulation pipeline to improve the authenticity of synthetic foggy images and construct a photo-realistic fog dataset to facilitate future research on defogging algorithms.

3. End-to-end Foggy Image Simulation

We propose an end-to-end approach for generating photo-realistic foggy images that incorporates accurate light transportation in scattering medium and physical characteristics of optics and sensor into the synthesized images, as illustrated in Fig. 4. We first introduce a foggy scene imaging model in Sec. 3.1. Based on this model, we establish a simulation pipeline consisting of two components: a) Foggy scene radiance is rendered using volumetric path tracing [38]. b) The radiance data is processed through a physically-based camera model, which comprises optics, sensor and image processing to faithfully replicate real camera devices [17]. These components are further detailed in Sec. 3.2.1 and Sec. 3.2.2, respectively.

3.1. Foggy Scene Imaging Model

The Atmospheric Scattering Model (ASM) [33, 34] has been widely used for the generation of foggy images [19,

25, 40, 53]:

$$E(d, \lambda) = \underbrace{E_o(\lambda)e^{-\beta(\lambda)d}}_{\text{transmission}} + \underbrace{L_h(\infty, \lambda)(1 - e^{-\beta(\lambda)d})}_{\text{airlight}}, \quad (1)$$

where E is the irradiance reaching the film under foggy conditions, E_o is the irradiance under clear conditions, β is the scattering coefficient characterizing the scatter ability of the medium, d is the distance between the object and the camera, $L_h(\infty, \lambda)$ denotes the radiance of airlight at infinity and λ refers to wavelength. Fig. 3a provides a more intuitive description of this process. Foggy images can be generated with pre-defined β and L_h values by directly using the clear images as E_o and the corresponding depth maps as d .

However, there are two issues with this scattering model for generating foggy images: (1) The definition of environmental illumination fails to consider the contribution of indirect illumination from the scene to the scattering medium, resulting in an unfaithful representation of the actual illumination. (2) The current formulation of airlight is insufficient for accurately modeling the lighting conditions that involve multiple point light sources, such as street lamps and automotive lights. This limitation restricts the model’s applicability in situations like nighttime [29, 51].

Furthermore, ASM only captures the irradiance reaching the sensor plane. Consequently, foggy images generated using ASM fail to accurately incorporate the realistic physical characteristics of optics and sensor in actual imaging process. In Fig. 2, we depict the relationship between noise level and fog density after the enhancement of regional contrast stretching [48] on a real captured image [8]. It can be seen that the noise level increases with higher fog density, emphasizing the critical need for accurate noise modeling in foggy image simulation.

To enhance the consistency between synthetic images and real-world captures, we develop a comprehensive foggy scene imaging model that incorporates the entire imaging process, referred to as:

$$I_{out} = f_{ISP}(f_{sensor}(f_{optics}(L_{scene})) + n), \quad (2)$$

where L_{scene} is the scene radiance obtained using volumetric path tracing [38], as described in Sec. 3.2.1; $\{f_{optics}, f_{sensor}, f_{ISP}\}$ denote the optics, sensor and ISP processes respectively, as described in Sec. 3.2.2, and the noise term introduced by the sensor is denoted as n .

By employing volumetric path tracing to obtain the radiance of foggy scene, we can precisely model the light scattering process under global illumination and active lighting. By incorporating the complete imaging process described in Eq. (2), we are able to faithfully replicate the authentic camera characteristics in foggy image simulation.

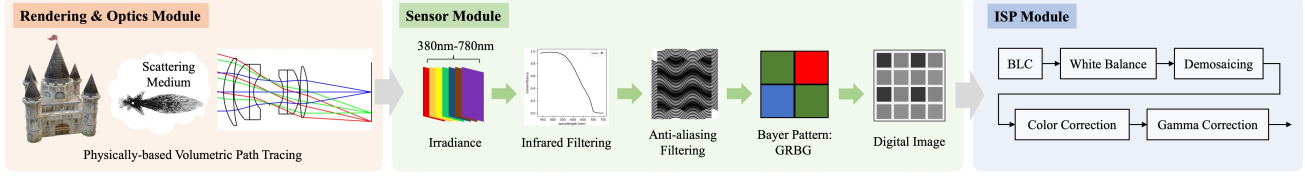


Figure 4. End-to-end foggy image simulation pipeline. The spectral radiance data is rendered using volumetric path tracing and passing through a realistic optics model before reaching the sensor plane. Subsequently, the irradiance is converted into a digital image through the sensor model. The raw image from the sensor is then processed by an ISP module to generate the final foggy image.

3.2. End-to-end Simulation Pipeline

3.2.1 Foggy Scene Rendering

We utilize procedural modeling techniques, similar to the approach employed in [31, 49], to create large-scale and diverse driving scenes. This involves defining adjustable parameters and using a set of predefined rules to generate explicit scene definitions. Our model library incorporates a diverse range of pedestrians, bicyclists and cars, each with well-defined material and texture properties. Furthermore, we carefully select 16 sky maps with cloudy conditions to ensure realistic lighting for foggy scenes, following the established guidelines from previous works [2–4].

In our approach, fog is treated as a spatial volume, and the spectral scene radiance is rendered using volumetric path tracing [38]. To accurately model the scattering process of the fog medium, we employ Henyey and Greenstein phase function [21, 38] to model its particle scattering characteristics, denoted as:

$$p_{\text{HG}}(\cos \theta) = \frac{1}{4\pi} \frac{1 - g^2}{(1 + g^2 + 2g(\cos \theta))^{3/2}}. \quad (3)$$

By empirically setting the asymmetry parameter $g = 0.87$, we can accurately simulate the probability distributions for scattering in a specific direction at a given point within the fog medium, as shown in Fig. 3b.

The rendering equation consists of two parts that account for reduction of radiance caused by out-scattering and increase of radiance due to in-scattering at point p in direction ω , represented by Eq. (4) and Eq. (5), respectively [38].

$$L_i(p', -\omega) = e^{-\int_0^d \sigma_s(p+t\omega, \omega) dt} L_o(p, \omega). \quad (4)$$

$$L_s(p, \omega) = L_e(p, \omega) + \sigma_s(p, \omega) \int_{S^2} p(p, \omega_i, \omega) L_i(p, \omega_i) d\omega_i. \quad (5)$$

σ_s represents the scattering probability per unit distance, $\{L_i, L_o, L_e, L_s\}$ are the {incident, exitant, emission, total added} radiance. In our implementation, we employ the Monte Carlo algorithm to solve the above integration.

The key distinction between our physically-based foggy scene rendering technique and ASM lies in the modeling of airlight, as represented by Eq. (5). Our approach offers a

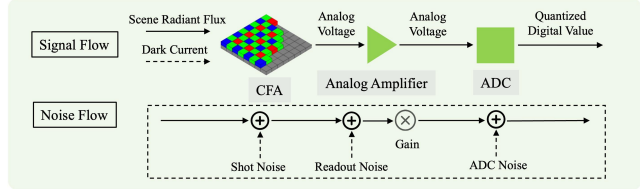


Figure 5. Noise flow and sensor simulation.

more accurate depiction of light scattering process through the utilization of specific phase function, and enables incorporation of global illumination and multiple light sources within the scene, producing authentic scene radiance to the camera model described in the next section.

3.2.2 Imaging through Realistic Camera Simulation

After obtaining the scene radiance, we employ ISETCam [15–17, 30, 32] to simulate the realistic digital camera system that encompasses optics, sensor and ISP components. Specifically, we use a diffraction-limited optics model [7] with consideration of off-axis irradiance attenuation based on the cosine-fourth law [24] to convert scene radiance L_o to optical image irradiance E_i , denoted as:

$$E_i = \pi \cdot L_o \frac{1}{1 + 4(f/\#(1 - m))^2} (\cos \phi)^4, \quad (6)$$

where $f/\#$ is the f-number, m represents the magnification defined as the negative ratio of image distance and object distance. In this optics model, blur mainly depends on the f-number value but little else.

We model the sensor process including photovoltaic conversion, analog amplifier, and analog-to-digital conversion as described in Fig. 4 and simulate the real noise characteristics depicted in Fig. 5. Among the various noise sources, shot noise arising from the photovoltaic conversion emerges as the dominant type of noise in foggy images which follows the Poisson distribution. To ensure the credibility of our sensor simulation, we incorporate a set of calibrated sensor parameters from the Mi10Pro cellphone, encompassing the CFA and infrared filter data.

Finally, we process the RAW image output from the sensor with a simple ISP module, including black level compensation, white balance, denoising, color correction and

Table 1. Comparisons between SynFog and existing synthetic fog datasets in the driving field.

Datasets	Global Illumination	Optics and Sensor Characteristics
Foggy Cityscapes	×	×
Virtual KITTI 2	×	×
SynFog (Ours)	✓	✓

gamma correction, to get the final synthetic image that mimics the real camera characteristics.

4. SynFog Dataset

By utilizing the proposed pipeline, we develop a new extensive synthetic fog dataset called SynFog. Example images are shown in Fig. 1. We first generate 500 large-scale, diverse driving scenes using procedural modeling techniques. For each scene, we generate two sets of images under different lighting conditions: one set with sky light only and the other set with both sky light and active light sources, such as street lamps and automotive lighting. Within each group, we create a set of four images: one clear image and three foggy images with varying fog levels defined by the scattering coefficient σ_s . According to McCartney’s work [33], the scattering coefficient remains independent of the wavelength for fog [13, 34]. Therefore, we use a consistent scattering coefficient for wavelengths within the visible range. To ensure realistic simulation of fog, we employ the meteorological optical range (MOR) [23] or visibility as a reference to select appropriate scattering coefficients, described as:

$$MOR = 2.996/\sigma_s. \quad (7)$$

Based on this equation, we select scattering coefficient values of 0.005, 0.01, 0.02, which correspond to visibility range of 600 m, 300 m and 150 m, respectively. According to National Standard of China [36], these visibility ranges are classified as heavy fog, thick fog, and dense fog, thereby the simulated foggy images are diversified enough to reflect real-world complicated foggy conditions.

We compare SynFog with existing fog datasets in the field of autonomous driving, including the extensively used Foggy Cityscapes [40] and Virtual KITTI 2 [18]. The comparison is presented in Tab. 1. In terms of foggy image generation, methods that rely on ASM [34] require accurate depth maps to calculate the correct transmission maps. However, obtaining accurate depth maps for outdoor images can be challenging, leading to artifacts such as holes and discontinuities in synthetic foggy images, similar to what is observed in Foggy Cityscapes [40]. On the other hand, foggy images generated using Game Engines like Unity or Unreal [18] typically employs rasterization techniques in the rendering process. While effective for creating

scenes, rasterization techniques may not accurately model light transportation as precisely as path tracing, which is employed in SynFog. Moreover, SynFog takes the realistic imaging process into account, resulting in highly authentic foggy images with realistic physical characteristics. The superior realism of SynFog is further supported by the experimental results.

Furthermore, our data simulation pipeline offers exceptional flexibility and control, allowing us to generate not only the final RGB images but also other valuable data modalities, including RAW images. This capability opens up avenues for research on defogging algorithms in the RAW domain, which has not been explored due to the lack of paired foggy RAW images and clear RGB images necessary for training deep defogging networks.

5. Experiments

5.1. Fog Chamber Validation

We establish an indoor fog chamber to validate the fidelity of our fog simulation pipeline. The chamber is constructed as a capped glass cube with dimensions of $1.5m \times 0.5m \times 0.5m$. To reproduce the dark night environment that has directional light sources in the fog chamber, we position 4 light bars on each side of the front face of the cuboid fog chamber and enclose it with black blackout cloth from the inside. Fog inside the chamber is generated using a 1.7MHz ceramic atomizer that emulates water mist. To ensure uniformity of fog concentration within the chamber, we incorporate dual fans to promote even distribution. A visibility measurement module is designed to quantify the fog density. This module is comprised of a PM-160 wireless power meter from Thorlabs, an 852 narrowband laser, an aperture and a narrowband filter adapted to the laser band. The aperture serves to filter out stray light and the narrowband filter plays the role in blocking interference caused by non-laser light sources in the scene. The principle of visibility (scattering coefficient σ_s) measurement is expressed as [37]:

$$\sigma_s = \frac{\ln\left(\frac{P_0}{P_u}\right)}{u}, \quad (8)$$

where P_0 is the laser power measured without fog and P_u is the measured laser power after the propagation of u distance under foggy conditions.

We then use Blender to create a virtual fog chamber that can reproduce the real-world counterpart. Our process involves calibrating the light conditions, object information and camera properties prior to capture, resulting in a one-to-one reproduction of the real fog chamber. We select the standard Macbeth color checker as our test object and use Mi10Pro cellphone to capture images. We evaluate the matching degree between the simulation and reality un-

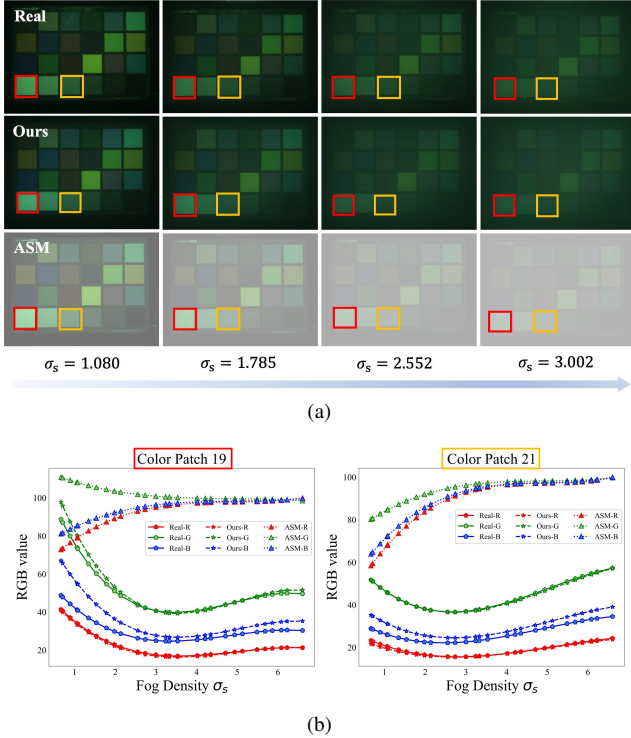


Figure 6. Color analysis under foggy conditions. (a) Visual comparison between real captures and simulated images under different fog densities in linear RGB space. We increase the brightness of the images ($\times 2$) for visual purpose. (b) Analyzing the trend of color variation across different fog densities.

der clear conditions by conducting a color analysis on the checker, which is available in the supplementary material.

After ensuring the consistency of initial capture conditions, we proceed to capture images under foggy conditions. We first record the laser power under clear condition as P_0 , and then proceed to generate fog until the object is no longer visible. During the fog dispersion process, we capture images at regular intervals of 1 second and record the laser energy P_u to obtain real-time fog density measurements. We utilize the measured fog density parameter σ_s to simulate corresponding foggy counterparts based on the proposed foggy image simulation pipeline.

In Fig. 6a, we showcase both the real-captured foggy images and simulated ones generated using our pipeline and the ASM [33]. It can be seen that images simulated through our pipeline exhibit a closer resemblance to real-captured foggy images. We also conduct a color analysis on 24 color patches under different fog densities as depicted in Fig. 6b. The results demonstrate that, as fog density increases, the trend of each channel in simulated images generated by our pipeline consistently follows that of the real captures. In contrast, the simulated images based on ASM do not exhibit this trend. These findings provide a quantitative validation of the realism of our foggy image simulation method.

5.2. Transferability across the Real-to-Virtual Gap

We evaluate 7 representative state-of-the-art defogging algorithms on SynFog dataset: DCP [20], CAP [57], Grid-DehazeNet [28], MSBDN [12], FFA-Net [39], AECR-Net [50], and DehazeFormer [43]. We directly apply the models trained on SynFog to real foggy images sourced from Foggy Zurich [11] and Seeing Through Fog [5] datasets, without undergoing any model fine-tuning or domain adaptation. During the training process, we adhere to the official settings of these methods, and all of them demonstrate convergence by the end of the training. The results are shown in Fig. 7. Despite DCP [20] and CAP [57] producing excessively dark outputs, models trained on SynFog can produce natural defogged images with minimal artifacts.

To underscore the superior realism of SynFog compared to existing synthetic fog datasets, we further enhance our evaluation by training the top-performing models (AECR-Net [50], DehazeFormer [43]) on two prominent synthetic fog datasets relevant to autonomous driving (Foggy Cityscapes [40], Virtual KITTI [18]). Subsequently, these models are evaluated on real fog datasets for a comparative analysis. Additionally, we utilize ASM [34] to create a SynFog- β dataset with clear images from SynFog and test it using DehazeFormer [43].

Comparison with existing synthetic fog datasets. To mitigate any potential impact arising from model training issues, we initially conduct tests for each model using the corresponding test set of the training data. Test results are available in the supplementary material. Both AECR-Net [50] and DehazeFormer [43] exhibit high performance, confirming the correctness of the model training process. Subsequently, the trained models are directly applied to real-world test foggy images. Results are shown in Fig. 9 and Tab. 2. Owing to the high authenticity of the SynFog dataset, models trained on it can effectively generalize to real foggy images and produce naturally colored defogged images. Conversely, models trained on the other two datasets display poorer generalization to real data, resulting in artifacts and color distortion in defogged images. We also evaluate the defogged images using a pre-trained YOLOv8 [22] model for detection, showcasing how SynFog enhances machine vision capabilities (See Tab. 3). More results can be found in the supplementary material.

It is noteworthy that the training set of Foggy Cityscapes [40] comprises 8925 data pairs, and Virtual KITTI [18] includes 3452 data pairs, both exceeding the 1350 data pairs in SynFog. However, models trained on SynFog demonstrate superior generalization capabilities on real-world data compared to them, highlighting the crucial role of data quality for researchers.

Comparison with SynFog- β synthesized by ASM. With the ability to obtain pixel-accurate depth information for each image, we can eliminate the impact of inaccurate

Table 2. Defogging evaluations with AECRNet on real-world fog datasets.

Training Set	O-Haze [2]			Foggy Zurich [11]	Foggy Driving [40]	BeDDE [56]
	PSNR \uparrow	SSIM \uparrow	DHQI [14] \uparrow	DHQI [14] \uparrow	DHQI [14] \uparrow	DHQI [14] \uparrow
Foggy Cityscapes	14.46	0.5737	43.40	52.06	51.55	36.07
Virtual KITTI	13.90	0.5315	42.80	50.94	47.46	33.42
SynFog	15.43	0.6116	44.46	54.16	52.07	43.28

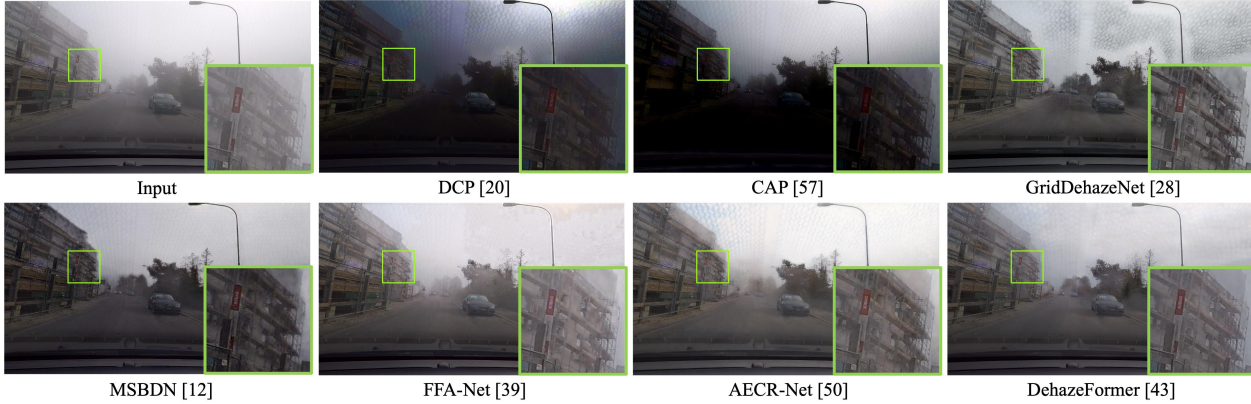


Figure 7. Transferability across real-world fog datasets. Test results of models trained on SynFog.



Figure 8. Ablation of sensor noise simulation.

Table 3. Detection evaluations on real-world fog datasets. “FZ” refers to Foggy Zurich. “STF” represents Seeing Through Fog.

Method	Training Set	FZ [11]	STF [5]
		mAP (%)	mAP (%)
AECRNet	Foggy Cityscapes	69.7	54.8
	Virtual KITTI	68.9	53.3
	SynFog	71.5	55.5
DehazeFormer	Foggy Cityscapes	67.9	54.9
	Virtual KITTI	68.5	53.1
	SynFog- β	59.7	55.3
	SynFog	69.7	55.3

depth and focus solely on the fog simulation method itself. The foggy image is synthesized from its clear counterpart in SynFog according to Eq. (1). Specifically, we employ the same airlight estimation method used in the construction of Foggy Cityscapes [40], which involves computing the median of all the 0.1% pixels with the largest dark channel values, as proposed in [45]. We use DehazeFormer [43] to train on both SynFog and SynFog- β , followed by testing the trained model on real-world foggy images. As shown in Fig. 9, image output by model trained on SynFog exhibits a better visual appearance compared to that of SynFog- β .

5.3. Ablation Study of Sensor Noise

To demonstrate the importance of sensor noise simulation in our pipeline, we create SynFog- α dataset. During the

Table 4. Impact of sensor noise on defogging performance.

Experimental setting	O-Haze [2]		
	PSNR \uparrow	SSIM \uparrow	DHQI \uparrow
AECRNet+SynFog(w/o noise)	15.16	0.5795	37.53
AECRNet+SynFog(w/ noise)	15.43	0.6116	44.46

Table 5. Impact of sensor noise on detection performance. “FZ” refers to Foggy Zurich. “STF” represents Seeing Through Fog.

Experimental setting	FZ [11]	STF [5]
	mAP (%)	mAP (%)
AECRNet+SynFog(w/o noise)	69.5	54.7
AECRNet+SynFog(w/ noise)	71.5	55.5

data generation process, the sensor-induced noise is set to zero, while all other settings remain consistent with those of the original SynFog dataset. We compare the defogging and detection performance of models trained on SynFog and SynFog- α datasets, as shown in Fig. 8 and Tab. 4-5. The model trained on SynFog that incorporates sensor noise produces more natural results with fewer artifacts and higher detection accuracy.

5.4. Algorithm Benchmarking

We use SynFog test set to assess the aforementioned representative defogging methods, employing both full-reference and no-reference metrics to evaluate the defogged results. To illustrate the challenges presented by SynFog, we also provide evaluation results from the widely-used RESIDE dataset [25]. The full-scale quantitative comparison results are available in the supplementary material.

Generally, learning-based methods consistently outperform earlier algorithms based on natural or statistical priors

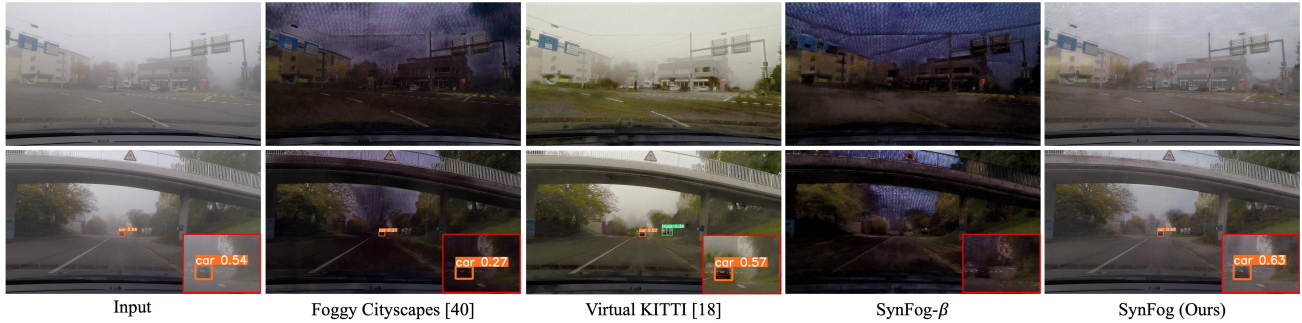


Figure 9. Visual and detection comparisons on real-world foggy images using DehazeFormer models trained on SynFog, SynFog- β , Foggy Cityscapes [40] and Virtual KITTI [18].



Figure 10. Qualitative comparisons of SOTA defogging methods on SynFog dataset with different fog levels. Zoom in for better view.

in terms of PSNR, SSIM and LPIPS [54]. In particular, DehazeFormer [43] attains the highest PSNR (27.96) and SSIM (0.923) values, while GridDehazeNet [28] achieves the best LPIPS (0.163) value on SynFog test set. Although the overall performance trend of the algorithms on SynFog is similar to that on RESIDE [25], the evaluation scores are noticeably lower, highlighting the challenges posed by SynFog. Fig. 10 depicts the qualitative comparison among these methods across different fog levels. Images restored by DCP [20] and CAP [57] tend to appear excessively dark and noisy due to inaccurate airlight estimation and the lack of effective denoising capabilities. GridDehazeNet [28] lacks color recovery ability in dense fog, while the outputs of FFA-Net [39] and DehazeFormer [43] seem smooth. This comparison further indicates the substantial room for improvement in existing defogging methods, especially when applied to more realistic fog datasets.

5.5. Discussion and Limitations

Experimental results reveal that models trained on SynFog, incorporating realistic physical characteristics of optics and sensors, exhibit a more robust performance on real-world fog datasets in terms of visual perception and detection accuracy. This outcome implies that integrating a physically-based end-to-end imaging process replicating real capture methods can significantly improve the generalization of

trained models. However, given the time-consuming nature of the physically-based volumetric path tracing technique, the scale of SynFog is not as extensive as other synthetic fog datasets. Nonetheless, we are committed to expanding the scene contents in the future and exploring a broader range of scattering mediums beyond fog.

6. Conclusion

In this paper, we present an end-to-end foggy image simulation pipeline. Our approach utilizes volumetric path tracing to model a more precise light scattering process with global illumination. By incorporating a physically-based camera processing pipeline that includes optics, sensor and image processing, we can closely mimic the authentic capture process under foggy conditions. Additionally, we develop a new synthetic fog dataset, SynFog, to facilitate the research on defogging. Comprehensive experiments have validated the authenticity and reliability of the SynFog dataset.

Acknowledgements This work is supported by 1) National Natural Science Foundation of China (Grant No. 62131011 and 62375233); 2) NSFC Young Scientists Fund (Grant No. 62302423). We thank Professor Chuxi Yang and Jiayue Xie for their help in preparing the scene creation. We also thank Professor Brian Wandell, Dr. Devesh Upadhyay, Dr. Alireza Rahimpour, and Devesh Shah for their valuable advice at the early stage of this project.

References

- [1] Cosmin Ancuti, Codruta O Ancuti, and Christophe De Vleeschouwer. D-hazy: A dataset to evaluate quantitatively dehazing algorithms. In *IEEE International Conference on Image Processing*, pages 2226–2230, 2016. **2**
- [2] Codruta O Ancuti, Cosmin Ancuti, Radu Timofte, and Christophe De Vleeschouwer. O-haze: a dehazing benchmark with real hazy and haze-free outdoor images. In *IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 754–762, 2018. **1, 2, 4, 7**
- [3] Codruta O Ancuti, Cosmin Ancuti, Mateu Sbert, and Radu Timofte. Dense-haze: A benchmark for image dehazing with dense-haze and haze-free images. In *IEEE International Conference on Image Processing*, pages 1014–1018, 2019.
- [4] Codruta O Ancuti, Cosmin Ancuti, and Radu Timofte. Nh-haze: An image dehazing benchmark with non-homogeneous hazy and haze-free images. In *IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 444–445, 2020. **1, 2, 4**
- [5] Mario Bijelic, Tobias Gruber, Fahim Mannan, Florian Kraus, Werner Ritter, Klaus Dietmayer, and Felix Heide. Seeing through fog without seeing fog: Deep multimodal sensor fusion in unseen adverse weather. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 11682–11692, 2020. **1, 3, 6, 7**
- [6] Norbert Buch, Sergio A Velastin, and James Orwell. A review of computer vision techniques for the analysis of urban traffic. *IEEE Transactions on Intelligent Transportation Systems*, 12(3):920–939, 2011. **1**
- [7] Ting Chen. *Digital camera system simulator and applications*. Stanford University, 2003. **4**
- [8] Lark Kwon Choi, Jaehee You, and Alan Conrad Bovik. Referenceless prediction of perceptual fog density and perceptual image defogging. *IEEE Transactions on Image Processing*, 24(11):3888–3901, 2015. **3**
- [9] Marius Cordts, Mohamed Omran, Sebastian Ramos, Timo Rehfeld, Markus Enzweiler, Rodrigo Benenson, Uwe Franke, Stefan Roth, and Bernt Schiele. The cityscapes dataset for semantic urban scene understanding. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 3213–3223, 2016. **2**
- [10] Dengxin Dai and Wen Yang. Satellite image classification via two-layer sparse coding with biased image representation. *IEEE Geoscience and Remote Sensing Letters*, 8(1):173–176, 2010. **1**
- [11] Dengxin Dai, Christos Sakaridis, Simon Hecker, and Luc Van Gool. Curriculum model adaptation with synthetic and real data for semantic foggy scene understanding. *International Journal of Computer Vision*, 128:1182–1204, 2020. **2, 6, 7**
- [12] Hang Dong, Jinshan Pan, Lei Xiang, Zhe Hu, Xinyi Zhang, Fei Wang, and Ming-Hsuan Yang. Multi-scale boosted dehazing network with dense feature fusion. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 2157–2167, 2020. **6**
- [13] J.B. Farrow et al. Influence of the atmosphere on optical systems. *Optica Acta*, 1970. **5**
- [14] X. Min et al. Objective quality evaluation of dehazed images. *TITS*, 2018. **7**
- [15] Joyce E Farrell and Brian A Wandell. Image systems simulation. *Handbook of Digital Imaging*, 1:373–400, 2015. **4**
- [16] Joyce E Farrell, Feng Xiao, Peter Bert Catrysse, and Brian A Wandell. A simulation tool for evaluating digital camera image quality. In *Image Quality and System Performance*, pages 124–131. SPIE, 2003.
- [17] Joyce E Farrell, Peter B Catrysse, and Brian A Wandell. Digital camera simulation. *Applied Optics*, 51(4):A80–A90, 2012. **3, 4**
- [18] Adrien Gaidon, Qiao Wang, Yohann Cabon, and Eleonora Vig. Virtual worlds as proxy for multi-object tracking analysis. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 4340–4349, 2016. **2, 5, 6, 8**
- [19] Martin Hahner, Dengxin Dai, Christos Sakaridis, Jan-Nico Zaech, and Luc Van Gool. Semantic understanding of foggy scenes with purely synthetic data. In *IEEE Intelligent Transportation Systems Conference (ITSC)*, pages 3675–3681, 2019. **2, 3**
- [20] Kaiming He, Jian Sun, and Xiaoou Tang. Single image haze removal using dark channel prior. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(12):2341–2353, 2010. **6, 8**
- [21] Louis G Henyey and Jesse Leonard Greenstein. Diffuse radiation in the galaxy. *Astrophysical Journal*, vol. 93, p. 70-83 (1941)., 93:70–83, 1941. **4**
- [22] Glenn Jocher, Ayush Chaurasia, and Jing Qiu. Ultralytics yolo, 2023. **6**
- [23] Kyung Won Kim. The comparison of visibility measurement between image-based visual range, human eye-based visual range, and meteorological optical range. *Atmospheric Environment*, 190:74–86, 2018. **5**
- [24] Miles V Klein and Thomas E Furtak. *Optik*. Springer-Verlag, 2013. **4**
- [25] Boyi Li, Wenqi Ren, Dengpan Fu, Dacheng Tao, Dan Feng, Wenjun Zeng, and Zhangyang Wang. Benchmarking single-image dehazing and beyond. *IEEE Transactions on Image Processing*, 28(1):492–505, 2018. **1, 2, 3, 7, 8**
- [26] Ruoteng Li, Xiaoyi Zhang, Shaodi You, and Yu Li. Realistic large-scale fine-depth dehazing dataset from 3d videos. *arXiv preprint arXiv:2004.08554*, 2020. **2**
- [27] Fayao Liu, Chunhua Shen, Guosheng Lin, and Ian Reid. Learning depth from single monocular images using deep convolutional neural fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 38(10):2024–2039, 2015. **2**
- [28] Xiaohong Liu, Yongrui Ma, Zhihao Shi, and Jun Chen. Grid-dehazenet: Attention-based multi-scale network for image dehazing. In *IEEE International Conference on Computer Vision*, pages 7314–7323, 2019. **6, 8**
- [29] Yun Liu, Zhongsheng Yan, Aimin Wu, Tian Ye, and Yuche Li. Nighttime image dehazing based on variational decomposition model. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 640–649, 2022. **3**

- [30] Zhenyi Liu, Trisha Lian, Joyce Farrell, and Brian Wandell. Soft prototyping camera designs for car detection based on a convolutional neural network. In *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops*, pages 0–0, 2019. 4
- [31] Zhenyi Liu, Minghao Shen, Jiaqi Zhang, Shuangting Liu, Henryk Blasinski, Trisha Lian, and Brian Wandell. A system for generating complex physically accurate sensor images for automotive applications. *arXiv preprint arXiv:1902.04258*, 2019. 4
- [32] Zheng Lyu, Thomas Goossens, Brian A Wandell, and Joyce Farrell. Validation of physics-based image systems simulation with 3-d scenes. *IEEE Sensors Journal*, 22(20):19400–19410, 2022. 4
- [33] Earl J McCartney. Optics of the atmosphere: scattering by molecules and particles. *New York*, 1976. 1, 3, 5, 6
- [34] Srinivasa G Narasimhan and Shree K Nayar. Vision and the atmosphere. *International Journal of Computer Vision*, 48:233–254, 2002. 3, 5, 6
- [35] Priya Narayanan, Xin Hu, Zhenyu Wu, Matthew D Thielke, John G Rogers, Andre V Harrison, John A D’Agostino, James D Brown, Long P Quang, James R Uplinger, et al. A multi-purpose realistic haze benchmark with quantifiable haze levels and ground truth. *IEEE Transactions on Image Processing*, 2023. 1, 2
- [36] National Standards of People’s Republic of China. Grade of fog forecast, 2011. 5
- [37] Roberto Nebuloni. Empirical relationships between extinction coefficient and visibility in fog. *Applied Optics*, 44(18):3795–3804, 2005. 5
- [38] Matt Pharr, Wenzel Jakob, and Greg Humphreys. *Physically based rendering: From theory to implementation*. MIT Press, 2023. 2, 3, 4
- [39] Xu Qin, Zhilin Wang, Yuanchao Bai, Xiaodong Xie, and Huizhu Jia. Ffa-net: Feature fusion attention network for single image dehazing. In *AAAI*, pages 11908–11915, 2020. 6, 8
- [40] Christos Sakaridis, Dengxin Dai, and Luc Van Gool. Semantic foggy scene understanding with synthetic data. *International Journal of Computer Vision*, 126:973–992, 2018. 1, 2, 3, 5, 6, 7, 8
- [41] Daniel Scharstein and Richard Szeliski. High-accuracy stereo depth maps using structured light. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages I–I. IEEE, 2003. 2
- [42] Nathan Silberman, Derek Hoiem, Pushmeet Kohli, and Rob Fergus. Indoor segmentation and support inference from rgb-d images. In *European Conference on Computer Vision*, pages 746–760. Springer, 2012. 2
- [43] Yuda Song, Zhuqing He, Hui Qian, and Xin Du. Vision transformers for single image dehazing. *IEEE Transactions on Image Processing*, 32:1927–1941, 2023. 6, 7, 8
- [44] Tao Sun, Mattia Segu, Janis Postels, Yuxuan Wang, Luc Van Gool, Bernt Schiele, Federico Tombari, and Fisher Yu. Shift: a synthetic driving dataset for continuous multi-task domain adaptation. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 21371–21382, 2022. 2
- [45] Ketan Tang, Jianchao Yang, and Jue Wang. Investigating haze-relevant features in a learning framework for image dehazing. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 2995–3000, 2014. 7
- [46] Jean-Philippe Tarel, Nicolas Hautiere, Aurélien Cord, Dominique Gruyer, and Houssam Halmaoui. Improved visibility of road scene images under heterogeneous fog. In *IEEE Intelligent Vehicles Symposium*, pages 478–485, 2010. 2
- [47] Jean-Philippe Tarel, Nicholas Hautiere, Laurent Caraffa, Aurélien Cord, Houssam Halmaoui, and Dominique Gruyer. Vision enhancement in homogeneous and heterogeneous fog. *IEEE Intelligent Transportation Systems Magazine*, 4(2):6–20, 2012. 2
- [48] Tali Treibitz and Yoav Y Schechner. Resolution loss without imaging blur. *JOSA A*, 29(8):1516–1528, 2012. 3
- [49] Apostolia Tsirikoglou, Joel Kronander, Magnus Wrenninge, and Jonas Unger. Procedural modeling and physically based rendering for synthetic data generation in automotive applications. *arXiv preprint arXiv:1710.06270*, 2017. 4
- [50] Haiyan Wu, Yanyun Qu, Shaohui Lin, Jian Zhou, Ruizhi Qiao, Zhizhong Zhang, Yuan Xie, and Lizhuang Ma. Contrastive learning for compact single image dehazing. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 10551–10560, 2021. 6
- [51] Jing Zhang, Yang Cao, Zheng-Jun Zha, and Dacheng Tao. Nighttime dehazing with a synthetic benchmark. In *Proceedings of the ACM International Conference on Multimedia*, pages 2355–2363, 2020. 3
- [52] Lin Zhang, Junhao Huang, Xiyuan Li, and Lu Xiong. Vision-based parking-slot detection: A dcnn-based approach and a large-scale benchmark dataset. *IEEE Transactions on Image Processing*, 27(11):5350–5364, 2018. 1
- [53] Lin Zhang, Anqi Zhu, Shiyu Zhao, and Yicong Zhou. Simulation of atmospheric visibility impairment. *IEEE Transactions on Image Processing*, 30:8713–8726, 2021. 2, 3
- [54] Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 586–595, 2018. 8
- [55] Yanfu Zhang, Li Ding, and Gaurav Sharma. Hazerd: an outdoor scene dataset and benchmark for single image dehazing. In *IEEE International Conference on Image Processing*, pages 3205–3209, 2017. 1, 2
- [56] Shiyu Zhao, Lin Zhang, Shuaiyi Huang, Ying Shen, and Shengjie Zhao. Dehazing evaluation: Real-world benchmark datasets, criteria, and baselines. *IEEE Transactions on Image Processing*, 29:6947–6962, 2020. 7
- [57] Qingsong Zhu, Jiaming Mai, and Ling Shao. A fast single image haze removal algorithm using color attenuation prior. *IEEE Transactions on Image Processing*, 24(11):3522–3533, 2015. 6, 8