

PhysAug: A Physical-guided and Frequency-based Data Augmentation for Single-Domain Generalized Object Detection

Xiaoran Xu^{1,2}, Jiangan Yang^{2*}, Wenhui Shi², Siyuan Ding², Luqing Luo², Jian Liu^{1,2*}

¹School of Advanced Interdisciplinary Sciences, University of Chinese Academy of Sciences

²IPHAC Lab, Institute of Microelectronics of the Chinese Academy of Sciences

xuxiaoran22@mails.ucas.ac.cn, {yangjiangan,luoluqing,shiwenhui,dingsiyuan,liujian}@ime.ac.cn

Abstract

Single-Domain Generalized Object Detection (S-DGOD) aims to train on a single source domain for robust performance across a variety of unseen target domains by taking advantage of an object detector. Existing S-DGOD approaches often rely on data augmentation strategies, including a composition of visual transformations, to enhance the detector’s generalization ability. However, the absence of real-world prior knowledge hinders data augmentation from contributing to the diversity of training data distributions. To address this issue, we propose PhysAug, a novel physical model-based non-ideal imaging condition data augmentation method, to enhance the adaptability of the S-DGOD tasks. Drawing upon the principles of atmospheric optics, we develop a universal perturbation model that serves as the foundation for our proposed PhysAug. Given that visual perturbations typically arise from the interaction of light with atmospheric particles, the image frequency spectrum is harnessed to simulate real-world variations during training. This approach fosters the detector to learn domain-invariant representations, thereby enhancing its ability to generalize across various settings. Without altering the network architecture or loss function, our approach significantly outperforms the state-of-the-art across various S-DGOD datasets. In particular, it achieves a substantial improvement of 7.3% and 7.2% over the baseline on DWD and Cityscape-C, highlighting its enhanced generalization in real-world settings.

Code — <https://github.com/startracker0/PhysAug>

Introduction

Object detection is a fundamental task in computer vision aiming to localize and recognize objects in natural images (Ren et al. 2015; Redmon et al. 2016; Tian et al. 2022; Zou et al. 2023). In past decades, object detection has achieved significant progress due to the advancement of deep learning, which assumes that the training data and test data come from the same distribution (Zhao et al. 2019). However, distribution shifts in various real-world applications make deep learning-based object detectors impractical. For instance, when trained on high-quality source data, object detectors for autonomous driving tend to perform poorly

on different distributions of target data due to the changes in weather, illuminations, and object appearances. Such performance degradation hinders its wide deployment in safety-critical areas.

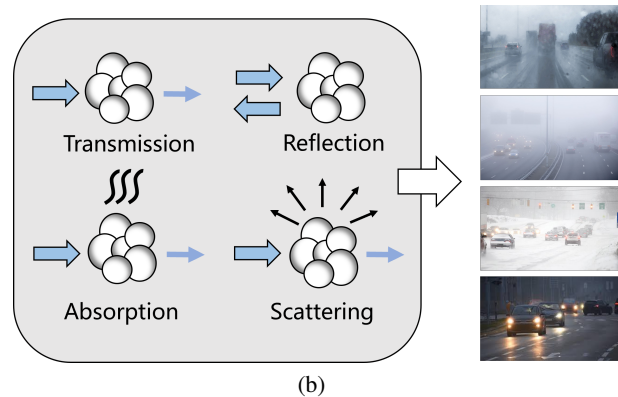
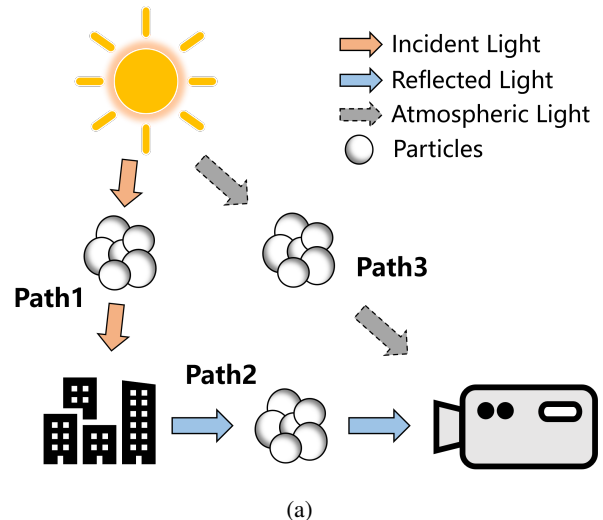


Figure 1: (a) The Physical imaging process in the atmosphere. Path1: Incident light path, Path2: Reflected light path, Path3: Atmospheric light path. (b) The cases of image degradation due to the different interactions between light and atmospheric particles.

*Corresponding Author.

To tackle this issue, single-domain generalized object detection (S-DGOD) has emerged as a prominent learning paradigm to alleviate the domain-shift impact, which aims at generalizing an object detector trained on a single source domain to multiple unseen target domains. Along this line of research, data augmentation is an active research direction for the S-DGOD problem to enrich the diversity of source domain. Several works propose to design the compositional policies of pre-defined visual corruptions applied on the global image or object-level regions (Lee et al. 2024) for expanding training samples. SRCD (Rao et al. 2023) introduces a mixing strategy of the global image with texture of its patches to enforce the model to focus on semantic content. UFR (Liu et al. 2024) adapts a combination of frequency-based and spatial-based transformations to augment local objects. However, these methods do not explicitly consider the formation mechanism of non-ideal imaging conditions, preventing the capture of real-world variations. Fig. 1(a) depicts the optical imaging process with three light propagation paths through the atmosphere. Among these paths, as shown in Fig. 1(b), interactions between light and various particles in the atmosphere are the primary causes of image degradation.

In this paper, we first formulate a universal perturbation model from the theory of atmospheric optics. This perturbation model aims to explicitly characterize the inferior cases of optical imaging resulting from light propagation properties in the atmosphere. Guided by this physical-based model, we introduce our data augmentation method, namely PhysAug. Specifically, PhysAug is implemented from a fourier perspective: we adopt simple frequency-dependent filters to augment low-frequency components of the image, simulating the global non-uniform illumination caused by light scattering and attenuation. Furthermore, fourier basis functions are randomly combined and projected onto the spatial domain of the image, generating local occlusion effects due to non-transparent particles in the wild. We conduct experiments on two S-DGOD datasets, Diverse Weather Dataset (DWD) and Cityscapes-C. The results show that PhysAug significantly outperforms existing state-of-the-art methods, achieving improvements of 7.3% and 7.2% over the baseline methods on the corrupted domains of DWD and Cityscapes-C, respectively. Besides, our analysis indicates that PhysAug enables models to learn better domain-invariant representations than non-physical based augmentations, proving its superiority. In summary, our contributions are as follows:

- We establish a universal perturbation model based on atmospheric optics, which is the first attempt to guide the design of data augmentation from the perspective of physical principles.
- We introduce PhysAug, a frequency-based data augmentation method tailored for the S-DGOD problem. This method accurately simulates naturally-occurring visual degradation throughout the full trajectory of light propagation, enriching real-world variations in training data.
- Comprehensive experiments and analysis validate the effectiveness of our method. Our method achieves new

state-of-the-art performance on two standard S-DGOD datasets, outperforming other approaches by a large margin.

Related Works

Single-Domain Generalization for Object Detection

To enhance model performance in unseen target domains, recent advancements in this field have explored various approaches. These approaches could be categorized into three main types: data augmentation, losses for domain generalization, and network architecture improvements. Most existing methods address this task through data augmentation. Those methods generate new domains via various transformations, thereby converting the S-DGOD problem into a domain adaptation object detection challenge. (Vidit, Engilberge, and Salzmann 2023) leveraged the vision-language model CLIP for semantic augmentation with textual prompts, enhancing the diversity of features extracted by the model. (Lee et al. 2024) employs an object-aware augmentation approach that enables the model to better detect and focus on potential regions of interest. DivAlign (Danish et al. 2024) similarly opts for diversity in the source domain and additionally aligns losses based on the class prediction results from the detector. In addition to data augmentation, domain-invariant feature distillation and neural architecture search have also been successfully employed to enhance model generalization. CSDS (Wu and Deng 2022) adopts a cyclic self-decoupling method to extract domain-invariant representations from the training data. G-NAS (Wu et al. 2024) searches for the most general architecture to prevent the model from overfitting. Unlike other methods, we propose PhysAug from the perspective of physical modeling, incorporating real-world prior knowledge into the augmentation process.

Data Augmentation for Domain Generalization

The most commonly used method in domain generalization is data augmentation in image domain (Zhou et al. 2022; Shorten and Khoshgoftaar 2019; Zhao et al. 2020; Rebuffi et al. 2021), there are several widely applied methods, including MixUp (Zhang et al. 2017), Cutout (DeVries and Taylor 2017), and AugMix (Hendrycks et al. 2019). Although image domain data augmentation has achieved significant success in enhancing model generalization and robustness. (Yin et al. 2019) found that models trained with visual transformations could be susceptible to noise affecting specific parts of the frequency spectrum. AmpMix (Xu et al. 2023) linearly interpolates the amplitudes of two images while keeping the phase information unchanged to enhance the model’s ability to generalize. Furthermore, some studies attempt to combine image domain and frequency domain augmentation methods. PRIME (Modas et al. 2022) employs a combined augmentation approach that integrates both frequency domain and image domain techniques. AFA (Vaish, Wang, and Strisciuglio 2024) builds upon the approach of PRIME and further expands the integration of image domain and frequency domain augmentation techniques.

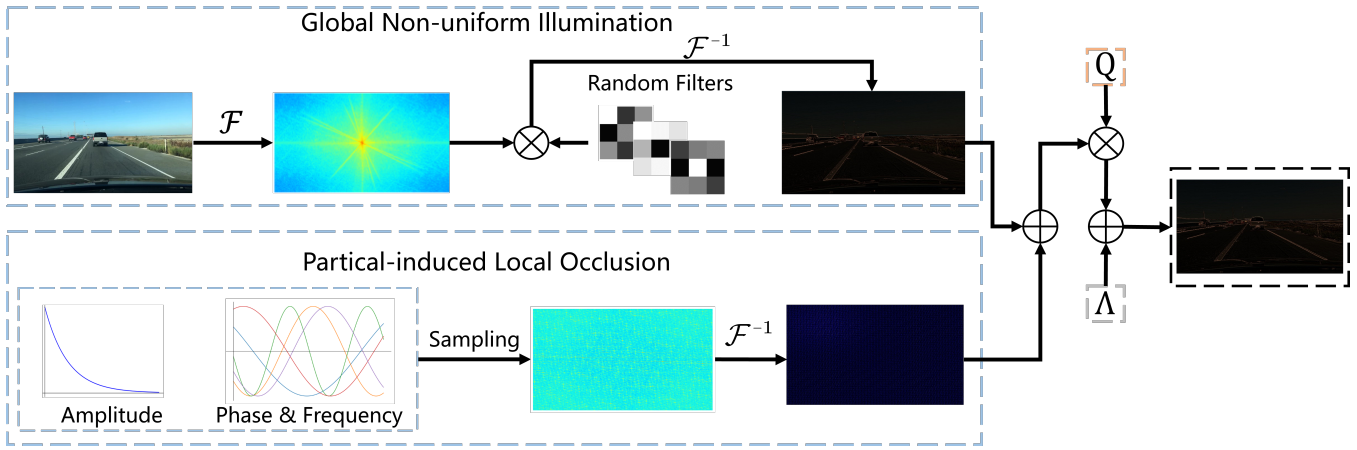


Figure 2: An overview of PhysAug. The input of PhysAug is an RGB training image, and its output is an augmented image. The two important components in PhysAug are Global Non-uniform Illumination (Top) and Particle-induced Local Occlusion (Bottom).

Preliminary

Problem Formulation

Single-domain generalization for object detection (S-DGOD) aims to train the model on a single source domain while ensuring it performs well across multiple target domains (Wang et al. 2021). Given a labeled source domain $D_s = \{(x_i^s, y_i^s)\}_{i=1}^{N_s}$ and multiple unseen target domains $M = \{D_t^1, D_t^2, \dots, D_t^k\}$, where $D_t^n = \{(x_i^t)\}_{i=1}^{N_t}$, N_s and N_t denote the size of the source and target datasets respectively. The S-DGOD task is to train an object detector F on the source domain D_s , and test its performance across target domains M .

Evaluation Metrics

We follow the settings established by (Michaelis et al. 2019), using mean average precision (mAP) as metrics to evaluate the performance of our method on the clean domain. Furthermore, we also use mean performance under corruption (mPC) established by (Lee et al. 2024) as metrics to evaluate our method’s performance on corrupted domains, which represents the average of mean average precision (mAP) across all corrupted domains and severity levels, the calculation is as follows.

$$mPC = \frac{1}{N_C} \sum_{C=1}^{N_C} \left(\frac{1}{N_S} \sum_{S=1}^{N_S} P_{C,S} \right), \quad (1)$$

$P_{C,S}$ is the target domain corrupted by corruption C at severity S , N_S and N_C means the number of severity and corruption. For DWD, the values for N_S, N_C are 1 and 4, for Cityscapes-C, the values are 5 and 15.

Methodology

In this section, we introduce the design and implementation of PhysAug. First, we formulate the universal perturbation model from the theory of atmosphere optics. Based on the model, we derive our data augmentation method from

a fourier perspective to explicitly simulate real-world variations for augmented data.

A Physical-based Perturbation Model

Physical imaging models based on atmospheric optics have achieved great success in various low-level computer vision tasks (Li, Cheong, and Tan 2019). Retinex theory (Land and McCann 1971) is introduced to simplify the optical imaging process as a combination of the incident light image $L(x)$ and the reflected light image $R(x)$, concentrating on the light propagation of Path1 and Path2 in Fig. 1(a). This can be represented by:

$$I(x, y) = L(x, y) \cdot R(x, y), \quad (2)$$

where $I(x)$ is the observed image, (x, y) is the image point. Retinex-based algorithms are designed to eliminate the effects of incident light to restore image quality, making them widely used for low-light enhancement tasks. The atmospheric scattering model (Nayar and Narasimhan 1999) is another popular approach to model physical imaging in foggy conditions. Its mathematical form is as follows:

$$I(x, y) = t(x, y) \cdot J(x, y) + A \cdot [1 - t(x, y)], \quad (3)$$

where $J(x, y)$ represents the haze-free image, A denotes the atmospheric light and $t(x, y)$ is the transmission function. The first term of Eq. (3) indicates light scattering effects along the reflected light path, corresponding to Path2 in Fig. 1(a). The second term refers to changes in atmospheric light path (e.g., Path3). The above modeling methods focus on special cases of the physical imaging process for image restoration. In practice, the quality of optical imaging is influenced by the interactions of light with different types of particles along all three propagation paths. So we formulate a perturbation model for covering general cases of inferior

imaging, which is defined as:

$$I(x, y) = Q \cdot [h_g(x, y) + h_o(x, y)] + A \cdot t_w(x, y), \quad (4)$$

$$h_g(x, y) = t_g(x, y) \cdot J(x, y), \quad (5)$$

$$h_o(x, y) = t_o(x, y) \cdot \sum_{i=0}^n P_i, \quad (6)$$

Let Q denotes the incident light, and A is the atmospheric light. P_i represents the set of non-transparent particles, where i indicates the particle type (e.g., dust and water droplets) with $i \in [0, 1, \dots, n]$. The transmission functions $t_g(x, y)$, $t_o(x, y)$ and $t_w(x, y)$ correspond to the reflected image J , the particle set P_i , and atmospheric light A , respectively. $h_g(x, y)$ represents the global non-uniform illumination on the image $J(x, y)$ due to light scattering and attenuation. $h_o(x, y)$ reflects local occlusion stemming from light reflection and absorption by non-transparent particles. These two components together account for the potential perturbations along the reflected light path. This model satisfies two key properties: (1) Completeness. All three light propagation paths are covered. (2) Heterogeneity. It considers interactions between light and a variety of particles. We use this model to guide the design of PhysAug.

The Design of PhysAug

Overview In this section, we introduce the implemental details of PhysAug. Based on Eq. (4), we define PhysAug as follows:

$$\text{PhysAug}(x) := Q \cdot (\hat{h}_g(x) + \hat{h}_o(x)) + A, \quad (7)$$

where $\hat{h}_g(\cdot)$ and $\hat{h}_o(\cdot)$ denote the estimates of $h_g(\cdot)$ and $h_o(\cdot)$, respectively. To simplify notation, we use A to replace $A \cdot t_w(x, y)$. Fig. 2 depicts the overall structure of PhysAug consisting of four components: The most significant components are non-uniform illumination h_g and local occlusions h_o along the reflected light path, as most visual degradation generates in this path. We adopt frequency-based pipelines to simulate these two components, as shown in Fig. 2. Besides, Q and A correspond to illumination changes from the incident light and atmospheric light path.

Global Non-uniform Illumination Recent studies have shown that illumination information exhibits excellent separation properties in the image frequency domain, and is mainly concentrated in the low-frequency component (Vaish, Wang, and Strisciuglio 2024). So we apply the convolution with random kernels to the image, adjusting the amplitude spectrum of the low-frequency components to generate a diverse range of non-uniform illumination conditions. This operation is defined as:

$$\begin{aligned} \hat{h}_g(x, y) &:= J(x, y) * G \\ &:= \mathcal{F}^{-1} \{ \mathcal{F} \{ J(x, y) \} \cdot \mathcal{F} \{ G \} \}. \end{aligned} \quad (8)$$

where \mathcal{F} and \mathcal{F}^{-1} represent the Fourier transform and the inverse Fourier transform, respectively. G denotes a convolutional filter with a kernel size of $n \times n$, and its parameters are drawn from a gaussian distribution. $*$ is the convolution operation.

Partical-induced Local Occlusion It is well known that visible light consists of components of different wavelengths. Here, we consider that local occlusions are generated from a composite of single-wavelength light which undergoes reflection and absorption by interacting with stmospheric particles. Inspired by previous fourier-based analysis (Vaish, Wang, and Strisciuglio 2024; Chen et al. 2021; Xu et al. 2023), we utilize the sinusoidal planar wave functions to simulate the occlusion case of single-wavelength light. The real part of planar wave function is defined as:

$$S^P(x, y) = \Re \left(C \cdot e^{j2\pi f \cdot T(\omega, \phi)} \right), \quad (9)$$

$$T(\omega, \phi) = x \cos(\omega) + y \sin(\omega) + \phi, \quad (10)$$

where C is a constraint to enforce a unit l2-norm of the planar wave. $\Re(\cdot)$ denotes the real part. $T(\cdot)$ is a regular function. f and ω represent the frequency and phase, respectively. ϕ is set to a constant of $\frac{\pi}{n}$, where $n \in \mathbb{N}$. P denotes light of the specific wavelength, and in practice, we set $P \in [R, G, B]$. Hence, we use $S(x, y) = [S^R(x, y), S^G(x, y), S^B(x, y)]$ to represent the operations on R,G,B channels. Finally, we introduce the composite function to construct local occlusions from multiple types of particles:

$$\hat{h}_o(x, y) = \sum_{i=0}^n M_i \cdot S_i(x, y), \quad (11)$$

where i is the particle type. M_i denotes a random matrix of the same size as the image for diversifying the inconsistent properties of local occlusions. According to Eq. (7), we implement PhysAug to increase the diversity of augmented data. Its parameter settings are provided in the implementation details.

Experiments

In this section, we first introduce the datasets and implement details. Then, we present our detailed experimental results on two standard S-DGOD datasets.

Datasets

We evaluate our approach on two large-scale object detection datasets, Diverse Weather Dataset (DWD) and Cityscapes-C, following standard S-DGOD settings. DWD is an urban-scene detection dataset, which consists of five weather conditions, i.e., Daytime Sunny, Night Sunny, Night Rainy, Dusk Rainy and Daytime Foggy (Wu and Deng 2022). This Dataset is collected from multiple autonomous driving benchmarks (Yu et al. 2020; Sakaridis, Dai, and Van Gool 2018; Hassaballah et al. 2020). Daytime Sunny serves as a source domain containing 18,205 images for training and 8,313 for evaluation. Night Sunny, Night Rainy, Dusk Rainy and Daytime Foggy are used as unseen target domains, with 26,158, 2,494, 3,501, and 3,775 images, respectively. Cityscapes-C is a robust detection benchmark, which synthesizes 15 types of common corruptions with five severity levels based on the validation set of Cityscapes (Michaelis et al. 2019). These corruptions are categorized into four groups: Noise, Blur, Digital, and

	Daytime Sunny	Night Sunny	Dusk Rainy	Night Rainy	Daytime Foggy	mPC
Baseline	50.4	37.5	29.2	14.6	33.1	30.2
IBN-Net	49.7	32.1	26.1	14.3	29.6	25.5
SW	50.6	33.4	26.3	13.7	30.8	26.1
IterNorm	43.9	29.6	22.8	12.6	28.4	23.4
ISW	51.3	33.2	25.9	14.1	31.8	26.3
SHADE	–	33.9	29.5	16.8	33.4	28.4
CDS	56.1	36.6	28.2	16.6	33.5	28.7
SRCD	–	36.7	28.8	17.0	35.9	29.6
Vidit et al.	51.3	36.9	32.3	18.7	38.5	31.6
OA-Mix	56.4	38.6	33.8	14.8	38.1	31.3
OA-DG	55.8	38.0	33.9	16.8	38.3	31.8
Div	50.6	39.4	37.0	22.0	35.6	33.5
DivAlign	52.8	<u>42.5</u>	<u>38.1</u>	24.1	37.2	<u>35.5</u>
UFR	<u>58.6</u>	40.8	33.2	19.2	<u>39.6</u>	33.2
PhysAug	60.2	44.9	41.2	<u>23.1</u>	40.8	37.5

Table 1: Comparison with state-of-the-art methods on the Diverse Weather Dataset (DWD). OA-Mix and Div are data augmentation methods in OA-DG and DivAlign, respectively. The mPC indicates the average performance across four adverse weather conditions. Bold numbers represent the highest performance in each column, and underlined numbers indicate the second-highest rank. The results for Daytime Sunny are reported in mAP.

Weather. We use the training set of Cityscapes as the source domain, and corrupted images are regarded as unseen target domains for validation.

Implementation Details

We first describe the implementation details of the proposed PhysAug. We set the filter G in Eq. (8) as the size of 3×3 with a gaussian distribution $\mathcal{N}(0, 4)$. The frequency f and phase ω in Eq. (9) are both drawn from a uniform distribution $(-512, 512)$, and ϕ is $\frac{-\pi}{4}$. The matrix M_i in Eq. (11) is sampled from 2D gaussian distribution $\mathcal{N}(0, \Sigma)$, where Σ is an identical matrix with the same size as the image. Q is set to 1 to indicate a constant incident light coefficient. Following the atmospheric scattering model, Λ represents the attenuation effects of atmospheric light, given by $\Lambda = L_\infty(1 - e^{-d})$, and L_∞ is the atmospheric light at infinity, which is set to 10^{-1} (Nayar and Narasimhan 1999). We sample d from a uniform distribution $(0, 10)$. For DWD dataset, we adopt the setting up in CDS (Wu and Deng 2022) and OA-DG (Lee et al. 2024). Specifically, we use Faster R-CNN (Ren et al. 2015) with a ResNet101 (He et al. 2016) backbone, and employ the SGD optimizer with a momentum of 0.9 and a weight decay of 0.0001. The learning rate is set to 0.001, and the batch size is 2. All other configurations align with those in the CDS and OA-DG. For Cityscapes-C, we use Faster R-CNN (Ren et al. 2015) with a ResNet50 (He et al. 2016) backbone and feature pyramid networks (FPN) (Lin et al. 2017) as the baseline model. The learning rate is set to 0.01, with a batch size of 8. The SGD optimizer is employed with a momentum of 0.9 and a weight decay of 0.0001. All experiments are conducted on NVIDIA RTX 3090 GPU.

Performance on Diverse Weather Conditions

We evaluate PhysAug and other the state-of-the-art methods on DWD dataset, including SW (Pan et al. 2019), IBN-Net (Pan et al. 2019), IterNorm (Huang et al. 2019), ISW (Choi et al. 2021), SHADE (Zhao et al. 2022), Clip the Gap (Vidit, Engilberge, and Salzmann 2023), OA-DG (Lee et al. 2024), DivAlign (Danish et al. 2024), and UFR (Liu et al. 2024). Note that Div and OA-Mix are data augmentation methods in DivAlign and OA-DG, respectively. Table 1 reports their performances on five different scenarios. We can see that PhysAug achieves the best detection accuracy of 60.2 mAP in Daytime Sunny, 44.9 mAP in Night Sunny, 41.2 mAP in Dusk Rainy, and 40.8 mAP in Daytime Foggy, respectively. Overall, our method obtains the largest improvement of 7.3 mPC compared with the baseline detector in adverse weather conditions.

Performance on Common Corruptions

To further prove the generality of our proposed PhysAug, we also compare PhysAug with popular data augmentation approaches on Cityscapes-C, including Cutout (DeVries and Taylor 2017), PhotoDistort (Redmon and Farhadi 2018), AutoAug (Zoph et al. 2020), AugMix (Hendrycks et al. 2019), StylizedAug (Geirhos et al. 2018). Besides, we select other three methods that combine data augmentation with loss function, SupCon (Khosla et al. 2020), FSCE (Sun et al. 2021) and OA-DG (Lee et al. 2024) for comparison. These results are summarized in Table 2. It is noted that our proposed PhysAug achieves the state-of-the-art performance, improving by an average of 7.2 mPC over the baseline detector. Besides, we can find that our method performs better on Noise and Blur corruption types, with the best result of 17.0 mAP on Shot Noise, and 25.5 mAP on Motion Blur, respectively.

Method	Clean	Noise				Blur			Weather				Digital			mPC	
		Gauss	Shot	Impulse	Defocus	Glass	Motion	Zoom	Snow	Frost	Fog	Bright	Contrast	Elastic	Pixel		JPEG
baseline	42.2	0.5	1.1	1.1	17.2	16.5	18.3	2.1	2.2	12.3	29.8	32.0	24.1	40.1	18.7	15.1	15.4
+Data augmentation																	
Cutout	42.5	0.6	1.2	1.2	17.8	15.9	18.9	2.0	2.5	13.6	29.8	32.3	24.6	<u>40.1</u>	18.9	15.6	15.7
PhotoDistort	42.7	1.6	2.7	1.9	17.9	14.1	18.7	2.0	2.4	16.5	36.0	39.1	27.1	39.7	18.0	16.4	16.9
AutoAug	42.4	0.9	1.6	0.9	16.8	14.4	18.9	2.0	1.9	16.0	32.9	35.2	26.3	39.4	17.9	11.6	15.8
AugMix	39.5	5.0	6.8	5.1	18.3	18.1	19.3	6.2	5.0	20.5	31.2	33.7	25.6	37.4	20.3	19.6	18.1
StylizedAug	36.3	4.8	6.8	4.3	19.5	18.7	18.5	2.7	3.5	17.0	30.5	31.9	22.7	33.9	22.6	<u>20.8</u>	17.2
OA-Mix	42.7	7.2	9.6	7.7	22.8	18.8	21.9	<u>5.4</u>	5.2	<u>23.6</u>	<u>37.3</u>	38.7	<u>31.9</u>	40.2	22.2	20.2	20.8
+Loss function																	
SupCon	43.2	7.0	9.5	7.4	22.6	20.2	<u>22.3</u>	4.3	5.3	23.0	<u>37.3</u>	38.9	31.6	<u>40.1</u>	24.0	20.1	20.9
FSCE	43.1	7.4	10.2	8.2	23.3	20.3	21.5	4.8	5.6	<u>23.6</u>	37.1	38.0	<u>31.9</u>	40.0	23.2	20.4	21.0
OA-DG	43.4	<u>8.2</u>	<u>10.6</u>	<u>8.4</u>	<u>24.6</u>	<u>20.5</u>	<u>22.3</u>	4.8	<u>6.1</u>	25.0	38.4	39.7	32.8	40.2	<u>23.8</u>	22.0	<u>21.8</u>
Ours																	
PhysAug	42.6	14.3	17.0	11.9	25.6	<u>19.1</u>	25.5	3.9	8.6	21.3	35.3	<u>39.5</u>	27.5	39.1	28.9	19.9	22.6

Table 2: The performance comparison with state-of-the-art methods on Cityscapes-C. The mPC is the average performance across 15 corruption types, each evaluated at 5 different severity levels. Numbers in bold represent the highest rank, and underlined numbers indicate the second highest rank.

Non-uniform Illumination	Local Occlusion	mAP	mPC
DWD			
✗	✗	50.4	30.2
✓	✗	60.4	36.5
✗	✓	60.6	35.2
✓	✓	60.2	37.5
Cityscapes-C			
✗	✗	42.2	15.4
✓	✗	42.4	20.5
✗	✓	43.0	20.2
✓	✓	42.6	22.6

Table 3: The performance comparison of each component in PhysAug on the DWD and Cityscapes-C datasets.

Discussion

Ablation Studies

In this section, we conduct ablation studies to analyze the efficacy of the proposed PhysAug and highlight the benefits of guidance from physical-based modeling.

PhysAug We now perform additional ablation experiments to study the impact of different components of the proposed PhysAug. Specifically, we compare the use of global non-uniform illumination in Eq. (8) and particle-induced local occlusions in Eq. (10). Table 3 reports on our results on DWD and Cityscapes-C dataset. For DWD dataset, each component can still obtain improvements of above 5 mPC in adverse weather conditions, indicating their versatility. The best performance is achieved with all components activated. One can see similar improvements of each component in Cityscape-C dataset, which validates the efficacy of the proposed method.

	Physical-based modeling	mAP	mPC
Baseline	-	42.2	15.4
NPM1	✗	42.5	22.2
NPM2	✗	42.2	19.4
PhysAug	✓	42.6	22.6

Table 4: The comparison of physical-based and non-physical-based modeling on the Cityscapes-C dataset.

Physical-based Modeling This experiment aims to validate the benefits of physical-based modeling. So we construct two non-physical models for comparison. These models are as follows:

$$I(x, y) = Q \cdot [h_g(x, y) + h_o(x, y)], \quad (12)$$

$$I(x, y) = \lambda \cdot Q \cdot [h_g(x, y) + h_o(x, y)] + (1 - \lambda) \cdot J(x, y) + L \quad (13)$$

where λ is the mixing factor. we name the models derived from Eq. (12) and Eq. (13) as NPM1 and NPM2. NPM1 neglects the negative effect of the atmospheric light path, and NPM2 follows the MixUp-style (Zhang et al. 2017) to mix the non-corruption image based on Eq. (4). We compare the performance of PhysAug with NPM1 and NPM2 on the Cityscapes-C dataset, as shown in Table 4. The results show that PhysAug achieves the largest performance gains on corrupted domains, with 0.4 – 3.2 mPC improvements on NPM1 and NPM2, respectively.

Qualitative Analysis

In this section, we further perform visualizations and feature assessment of our method to see how exactly it works.

Visualization Analysis In Fig. 3, we provide visualized object detection examples of the baseline method, OA-Mix and the proposed PhysAug on four adverse weather

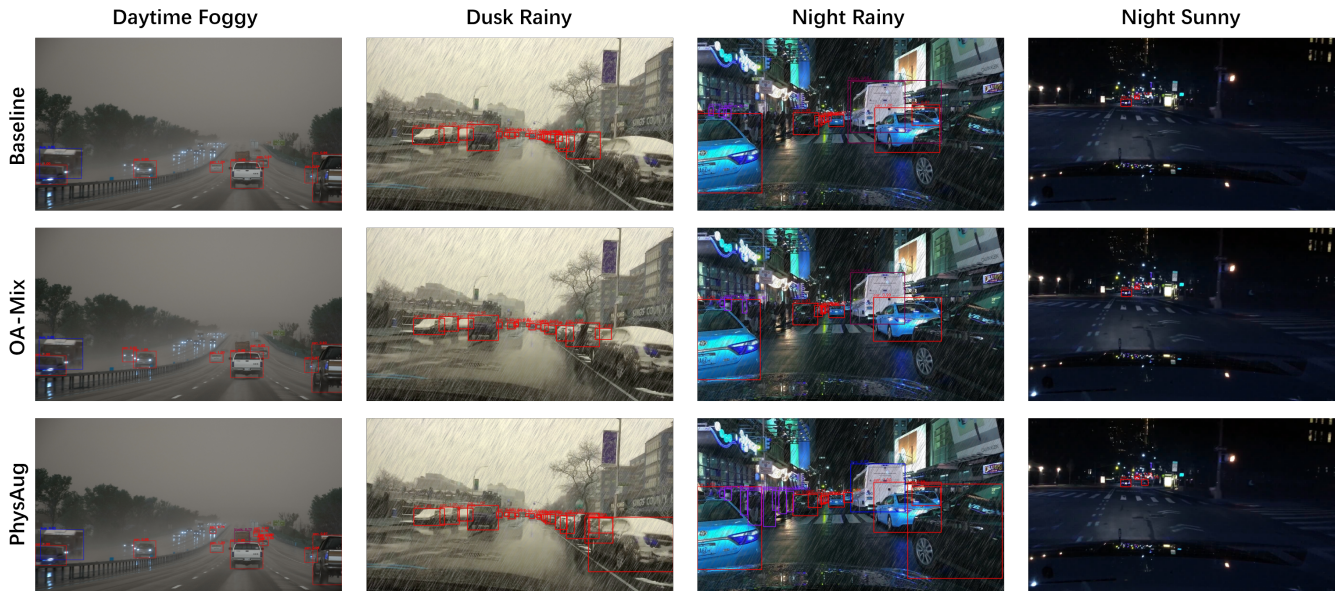


Figure 3: Visualized detection samples of the baseline method, OA-Mix and PhysAug in different weather conditions.

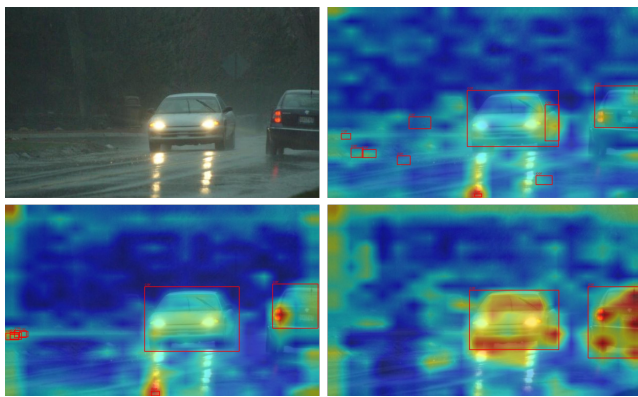


Figure 4: Heatmaps comparison in a night-rainy scenario. Top-left: original image. Top-right: The baseline method. Bottom-left: OA-Mix. Bottom-right: PhysAug.

conditions. We can see that compared with the baseline method and OA-Mix, PhysAug can largely reduce the rate of false and missing detection in the image. Although night rainy weather is full of challenge, our method can localize and recognize these small and obscure persons accurately, which further demonstrates the effectiveness of the proposed PhysAug.

Feature Analysis The heatmaps in Fig. 4 highlight the effectiveness of PhysAug in comparison to baseline and OAMix methods under challenging nighttime rainy conditions. The baseline method shows weak focus and sparse heatmap activation, indicating a limited ability to localize key objects such as vehicles in poor-visibility environments. OAMix improves target detection with more concentrated heatmap regions, particularly around the primary ve-

hicle, but still leaves peripheral areas underrepresented. In contrast, PhysAug demonstrates significantly enhanced detection performance, with highly concentrated and precise heatmap activations around the vehicles under adverse visual conditions. These results suggest that PhysAug effectively captures spatial and frequency-based features, resulting in superior object localization and robustness.

Conclusion

In this paper, we propose PhysAug, a novel data augmentation method for single-domain generalized object detection. To achieve this goal, we introduce a universal perturbation model from the theory of atmospheric optics to describe the general case of the inferior imaging process. Guided by this model, PhysAug simulates real-world variations caused by atmospheric particles to diversify training data. We conduct extensive qualitative and quantitative experiments to demonstrate the effectiveness of our method for the S-DGOD task. Results also show that PhysAug outperforms the state-of-the-art by a large margin.

Acknowledgements

This work was supported by the National Key Research and Development Program of China under Grant 2021YFB2501403.

References

Chen, G.; Peng, P.; Ma, L.; Li, J.; Du, L.; and Tian, Y. 2021. Amplitude-phase recombination: Rethinking robustness of convolutional neural networks in frequency domain. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 458–467.

- Choi, S.; Jung, S.; Yun, H.; Kim, J. T.; Kim, S.; and Choo, J. 2021. Robustnet: Improving domain generalization in urban-scene segmentation via instance selective whitening. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 11580–11590.
- Danish, M. S.; Khan, M. H.; Munir, M. A.; Sarfraz, M. S.; and Ali, M. 2024. Improving Single Domain-Generalized Object Detection: A Focus on Diversification and Alignment. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 17732–17742.
- DeVries, T.; and Taylor, G. W. 2017. Improved regularization of convolutional neural networks with cutout. *arXiv preprint arXiv:1708.04552*.
- Geirhos, R.; Rubisch, P.; Michaelis, C.; Bethge, M.; Wichmann, F. A.; and Brendel, W. 2018. ImageNet-trained CNNs are biased towards texture; increasing shape bias improves accuracy and robustness. *arXiv preprint arXiv:1811.12231*.
- Hassaballah, M.; Kenk, M. A.; Muhammad, K.; and Minaee, S. 2020. Vehicle detection and tracking in adverse weather using a deep learning framework. *IEEE transactions on intelligent transportation systems*, 22(7): 4230–4242.
- He, K.; Zhang, X.; Ren, S.; and Sun, J. 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 770–778.
- Hendrycks, D.; Mu, N.; Cubuk, E. D.; Zoph, B.; Gilmer, J.; and Lakshminarayanan, B. 2019. Augmix: A simple data processing method to improve robustness and uncertainty. *arXiv preprint arXiv:1912.02781*.
- Huang, L.; Zhou, Y.; Zhu, F.; Liu, L.; and Shao, L. 2019. Iterative normalization: Beyond standardization towards efficient whitening. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 4874–4883.
- Khosla, P.; Teterwak, P.; Wang, C.; Sarna, A.; Tian, Y.; Isola, P.; Maschinot, A.; Liu, C.; and Krishnan, D. 2020. Supervised contrastive learning. *Advances in neural information processing systems*, 33: 18661–18673.
- Land, E. H.; and McCann, J. J. 1971. Lightness and retinex theory. *Josa*, 61(1): 1–11.
- Lee, W.; Hong, D.; Lim, H.; and Myung, H. 2024. Object-Aware Domain Generalization for Object Detection. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, 2947–2955.
- Li, R.; Cheong, L.-F.; and Tan, R. T. 2019. Heavy rain image restoration: Integrating physics model and conditional adversarial learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 1633–1642.
- Lin, T.-Y.; Dollár, P.; Girshick, R.; He, K.; Hariharan, B.; and Belongie, S. 2017. Feature pyramid networks for object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2117–2125.
- Liu, Y.; Zhou, S.; Liu, X.; Hao, C.; Fan, B.; and Tian, J. 2024. Unbiased Faster R-CNN for Single-source Domain Generalized Object Detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 28838–28847.
- Michaelis, C.; Mitzkus, B.; Geirhos, R.; Rusak, E.; Bringmann, O.; Ecker, A. S.; Bethge, M.; and Brendel, W. 2019. Benchmarking robustness in object detection: Autonomous driving when winter is coming. *arXiv preprint arXiv:1907.07484*.
- Modas, A.; Rade, R.; Ortiz-Jiménez, G.; Moosavi-Dezfooli, S.-M.; and Frossard, P. 2022. Prime: A few primitives can boost robustness to common corruptions. In *European Conference on Computer Vision*, 623–640. Springer.
- Nayar, S. K.; and Narasimhan, S. G. 1999. Vision in bad weather. In *Proceedings of the seventh IEEE international conference on computer vision*, volume 2, 820–827. IEEE.
- Pan, X.; Zhan, X.; Shi, J.; Tang, X.; and Luo, P. 2019. Switchable whitening for deep representation learning. In *Proceedings of the IEEE/CVF international conference on computer vision*, 1863–1871.
- Rao, Z.; Guo, J.; Tang, L.; Huang, Y.; Ding, X.; and Guo, S. 2023. Srcd: Semantic reasoning with compound domains for single-domain generalized object detection. *arXiv preprint arXiv:2307.01750*.
- Rebuffi, S.-A.; Goyal, S.; Calian, D. A.; Stimberg, F.; Wiles, O.; and Mann, T. A. 2021. Data augmentation can improve robustness. *Advances in Neural Information Processing Systems*, 34: 29935–29948.
- Redmon, J.; Divvala, S.; Girshick, R.; and Farhadi, A. 2016. You only look once: Unified, real-time object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 779–788.
- Redmon, J.; and Farhadi, A. 2018. Yolov3: An incremental improvement. *arXiv preprint arXiv:1804.02767*.
- Ren, S.; He, K.; Girshick, R.; and Sun, J. 2015. Faster r-cnn: Towards real-time object detection with region proposal networks. *Advances in neural information processing systems*, 28.
- Sakaridis, C.; Dai, D.; and Van Gool, L. 2018. Semantic foggy scene understanding with synthetic data. *International Journal of Computer Vision*, 126: 973–992.
- Shorten, C.; and Khoshgoftaar, T. M. 2019. A survey on image data augmentation for deep learning. *Journal of big data*, 6(1): 1–48.
- Sun, B.; Li, B.; Cai, S.; Yuan, Y.; and Zhang, C. 2021. Fscc: Few-shot object detection via contrastive proposal encoding. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 7352–7362.
- Tian, Z.; Chu, X.; Wang, X.; Wei, X.; and Shen, C. 2022. Fully convolutional one-stage 3d object detection on lidar range images. *Advances in Neural Information Processing Systems*, 35: 34899–34911.
- Vaish, P.; Wang, S.; and Strisciuglio, N. 2024. Fourier-basis functions to bridge augmentation gap: Rethinking frequency augmentation in image classification. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 17763–17772.

Vidit, V.; Engilberge, M.; and Salzmann, M. 2023. Clip the gap: A single domain generalization approach for object detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 3219–3229.

Wang, X.; Huang, T. E.; Liu, B.; Yu, F.; Wang, X.; Gonzalez, J. E.; and Darrell, T. 2021. Robust object detection via instance-level temporal cycle confusion. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 9143–9152.

Wu, A.; and Deng, C. 2022. Single-domain generalized object detection in urban scene via cyclic-disentangled self-distillation. In *Proceedings of the IEEE/CVF Conference on computer vision and pattern recognition*, 847–856.

Wu, F.; Gao, J.; Hong, L.; Wang, X.; Zhou, C.; and Ye, N. 2024. G-NAS: Generalizable Neural Architecture Search for Single Domain Generalization Object Detection. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, 5958–5966.

Xu, Q.; Zhang, R.; Fan, Z.; Wang, Y.; Wu, Y.-Y.; and Zhang, Y. 2023. Fourier-based augmentation with applications to domain generalization. *Pattern Recognition*, 139: 109474.

Yin, D.; Gontijo Lopes, R.; Shlens, J.; Cubuk, E. D.; and Gilmer, J. 2019. A fourier perspective on model robustness in computer vision. *Advances in Neural Information Processing Systems*, 32.

Yu, F.; Chen, H.; Wang, X.; Xian, W.; Chen, Y.; Liu, F.; Madhavan, V.; and Darrell, T. 2020. Bdd100k: A diverse driving dataset for heterogeneous multitask learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2636–2645.

Zhang, H.; Cisse, M.; Dauphin, Y. N.; and Lopez-Paz, D. 2017. mixup: Beyond empirical risk minimization. *arXiv preprint arXiv:1710.09412*.

Zhao, L.; Liu, T.; Peng, X.; and Metaxas, D. 2020. Maximum-entropy adversarial data augmentation for improved generalization and robustness. *Advances in Neural Information Processing Systems*, 33: 14435–14447.

Zhao, Y.; Zhong, Z.; Zhao, N.; Sebe, N.; and Lee, G. H. 2022. Style-hallucinated dual consistency learning for domain generalized semantic segmentation. In *European conference on computer vision*, 535–552. Springer.

Zhao, Z.-Q.; Zheng, P.; Xu, S.-T.; and Wu, X. 2019. Object Detection With Deep Learning: A Review. *IEEE Transactions on Neural Networks and Learning Systems*, 30(11): 3212–3232.

Zhou, K.; Liu, Z.; Qiao, Y.; Xiang, T.; and Loy, C. C. 2022. Domain generalization: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(4): 4396–4415.

Zoph, B.; Cubuk, E. D.; Ghiasi, G.; Lin, T.-Y.; Shlens, J.; and Le, Q. V. 2020. Learning data augmentation strategies for object detection. In *Computer Vision—ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXVII 16*, 566–583. Springer.

Zou, Z.; Chen, K.; Shi, Z.; Guo, Y.; and Ye, J. 2023. Object detection in 20 years: A survey. *Proceedings of the IEEE*, 111(3): 257–276.