Robust Object Detection in Challenging Weather Conditions

Himanshu Gupta
AASS
Örebro University, Sweden
himanshu.gupta@oru.se

Oleksandr Kotlyar
AASS
Örebro University, Sweden
oleksandr.kotlyar@oru.se

Henrik Andreasson
AASS
Örebro University, Sweden
henrik.andreasson@oru.se

Achim J. Lilienthal
Perception for Intelligent Systems
TUM, Germany
achim.lilienthal@tum.de

Abstract

Object detection is crucial in diverse autonomous systems like surveillance, autonomous driving, and driver assistance, ensuring safety by recognizing pedestrians, vehicles, traffic lights, and signs. However, adverse weather conditions such as snow, fog, and rain pose a challenge, affecting detection accuracy and risking accidents and damage. This clearly demonstrates the need for robust object detection solutions that work in all weather conditions. We employed three strategies to enhance deep learning-based object detection in adverse weather: training on real-world all-weather images, training on images with synthetic augmented weather noise, and integrating object detection with adverse weather image denoising. The synthetic weather noise is generated using analytical methods, GAN networks, and style-transfer networks. We compared the performance of these strategies by training object detection models using real-world all-weather images from the BDD100K dataset and for assessment employed unseen real-world adverse weather images. Adverse weather denoising methods were evaluated by denoising real-world adverse weather images and the results of object detection on denoised and original noisy images were compared. We found that the model trained using all-weather real-world images performed best, while the strategy of doing object detection on denoised images performed worst.

1. Introduction

Object detection is an essential component in autonomous driving, ensuring the identification of pedestrians, vehicles, traffic signals, and obstacles to enhance safety. The advancement of the deep-learning approach and the availability of diverse data have led to robust and accurate models for object detection [18]. However, object detection in challenging weather conditions like snow, fog, and rain results in reduced accuracy due to obscurity/absence of salient object features, noise from weather patterns (rain streaks and snowflakes), lens interference, and decreased ambient light.

Efforts to address the issue of robust object detection in adverse weather have resulted in several strategies to improve object detection accuracy. One such strategy involves training deep learning models with well-annotated real-world datasets encompassing all weather conditions. While datasets such as [38], [2], and [19] offer images depicting diverse weather conditions, they often lack well-balanced all-weather/daylight conditions or do not have object annotations for adverse weather images. As these datasets are not holistic, the second option is to augment clear weather images through physics-based rendering approaches [40], [8], or generative adversarial networks [13], [21], or a fusion of both as demonstrated in [33]. Each method carries distinct advantages and limitations, with GANs achieving complex noise patterns at the cost of potentially altering image content drastically. At the same time, physics-based approaches lack realism in noise patterns but maintain image integrity. The third option is to do denoising and then detect objects for enhanced accuracy. Several image-denoising methods exist that focus on dehazing [32] [10], deraining [20], and desnowing [17]. However, most denoising methods are evaluated based on image quality improvement rather than object detection performance on denoised images.

Given these strategies' varied nature and potential implications, a systematic and comparative evaluation of each approach becomes imperative. By meticulously assessing the strengths and weaknesses of these strategies, we can

work toward more robust, accurate, and reliable object detection systems that operate in all weather conditions. Consequently, this work evaluates the first and second strategies by training YOLOv5 [14] models using the BDD100k [38] dataset and testing their performance in unseen all-weather conditions from the DAWN [19] dataset. The third strategy is examined by selecting various analytical and deep learning-based adverse weather denoising methods, denoising real-world adverse weather images, followed by object detection assessment. During this study, we tried to answer these three questions: Can simple image augmentations like blurring, noising, and occlusion be enough to improve object detection in adverse weather using clear weather images? Are synthetic weather augmentation (analytical, GAN, or style transfer) in the absence of real all-weather images helpful in training a robust object detection model? Lastly, are current image-denoising methods sufficient for robust object detection in adverse weather? That leads to the following contributions of this work.

1. Evaluation of image augmentation strategies for robust object detection in adverse weather.
2. Synthetic weather noise generation using style-transfer network and impact on object detection accuracy under adverse weather conditions.
3. Evaluation of adverse weather image denoising in tandem with object detection.

2. Literature Review

The performance of object detection algorithms in adverse weather conditions is challenging due to reduced visibility, lighting variations, and weather-induced noise. These challenges are addressed in literature by improving weather augmentation techniques, varying training approaches, and creating denoising techniques to improve image quality and enhance object detection models' robustness and accuracy in adverse weather conditions.

Several analytical methods have been proposed in the literature for creating realistic weather augmentation in clear weather images. For instance, fog effect generation often employs Beer-Lambert's law of attenuation for a light beam passing through particles, as demonstrated in [40] [30] with diverse ways to estimate the depth map. In the context of rain rendering, [12] combines clear images with rain and fog layers based on scene depth. Meanwhile, [33] utilizes a physics-based particle simulator for individual rain streaks generation as a function of rainfall rate. For snow augmentation, [34] generated a 3D scene with snow using snowflake density, snowflake sizes, and relative velocity in OpenGL and rendered a 2D snow layer which is blended with a clear weather image. However, these methods commonly overlook image illumination while generating adverse weather, focusing more on denoising applications than object detection, which additionally requires label verification. While analytical methods fall short in replicating intricate noise patterns, GAN-based approaches, explored in [13] [21], are more effective in generating complex noise but fail to generate rainstreaks and snowfall. Consequently, [33] introduces a fusion of physics-based rain rendering with a GAN-based approach.

In addition to realistic weather augmentation, innovative training methods for object detection in adverse weather have also been researched. Some incorporate end-to-end training of denoising GANs and object detection models to enhance denoising and object detection as explored in [1] [31]. However, GANs may introduce spurious details, potentially affecting smaller objects like pedestrians. Other work like [25] introduced training a convolution network to generate parameters for effective dehazing using the analytical method along with the object detection pipeline. These methods predominantly pertain to foggy weather conditions, and their general applicability still needs to be explored.

Combining adverse weather denoising methods with object detection holds promise for improved performance in adverse conditions. However, the results of [26] and [20] showed degradation with this approach. [26] concludes that dehazing algorithms have minimal impact on heavy fog but are effective for images with moderate fog intensity. Furthermore, [20] evaluates deraining methods and observes degraded object detection performance (mean average precision) on derained images in driving scenarios.

Most of these studies undergo separate evaluations and are contrasted against baseline models without retraining for specialized tasks like autonomous driving. This highlights the importance of our study, which comprehensively evaluates the effectiveness of different strategies to enhance object detection's robustness, contributing to the progress of robust object detection in adverse weather conditions.

3. Methods and Material

This section presents a comprehensive overview of the technical approaches used in this study to generate synthetic weather noise, detailing the techniques employed and the models selected to address the challenges of enhancing object detection accuracy in adverse weather conditions.

3.1. Adverse Weather Augmentation

Several strategies exist for creating adverse weather noise from analytical or deep learning-based approaches. This study employs three simplified approaches for weather noise generation: analytical approaches, GAN networks, and style-transfer networks, and the code[1] is made avail-

[1] https://github.com/hgupta01/Weather_Effect_Generator

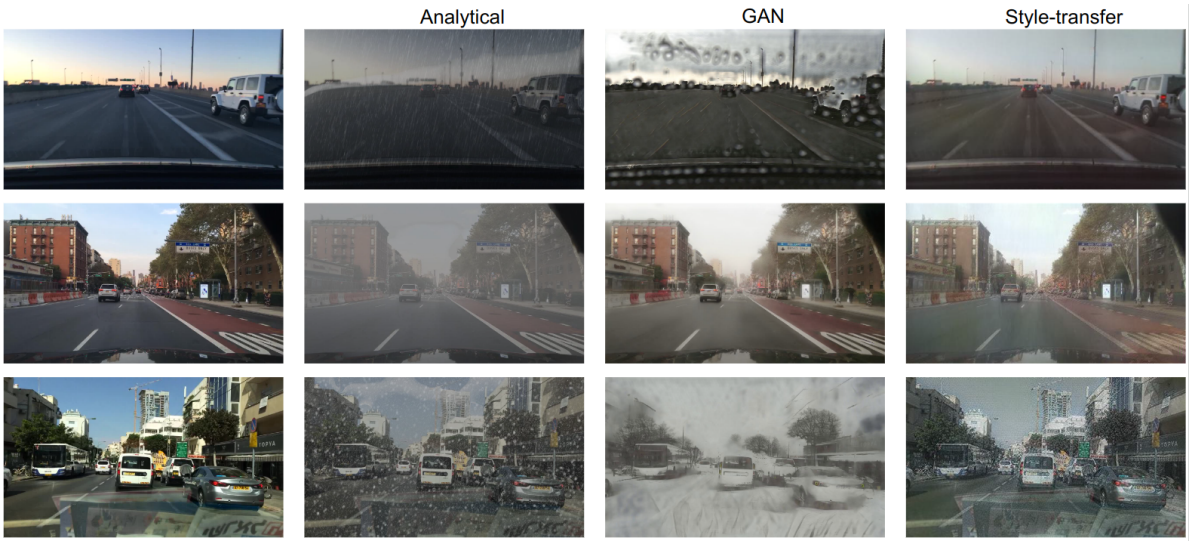

Figure 1. Weather augmentation using physics-based method (2nd column), GAN-based network (3rd column), and style-transfer network (4th columns) using the clear weather images (1st column) for rain (1st row), fog (2nd row), and snow (3rd row).

able for reproducibility. Fig 1 shows the augmented images using the methodologies implemented in this work. The augmentation techniques are employed in an offline manner, implying that chosen images are first augmented and then annotations are checked manually.

3.1.1 Analytical Method of Noising

The analytical approach implemented in our work utilizes the illumination information for deciding the fog color, visibility distance, image darkening factor, and the illumination threshold value (t_i) used to calculate the alpha channel (α) for blending the rain or snow layer with a clear weather image. There are a few common steps/functions employed in this work for generating the fog, rain, and snow augmentation which are as follows:

- Illumination estimation: The LIME method [11] is used to estimate an illumination map by finding the maximum values across RGB channels and refining the obtained map by imposing a structure prior. The illumination map is a 2D array with values in the range of $0, 1$ which is used to calculate the histogram with four bins representing the illumination classes (dark, low brightness, moderate brightness, and bright). The histogram is normalized and the illumination threshold (t_i) is the maximum value.

- Image darkening: The image brightness is reduced by multiplying the saturation and value channels of HSV images with a factor of 0.75, 0.65, and 0.5 for less bright, moderately bright, and bright images respectively, followed by conversion to RGB colorspace.

- Depth map (D) estimation: MiDaS 3.0 DPTL-384 [28] is utilized to estimate depth maps for clear weather images.

- Alpha channel (α) estimation: For alpha channel estimation, images are converted to greyscale with values in the range of $0, 1$, then the pixel values (p) greater than t_i are replaced by $1 - p$, followed by blur using a kernel of size 11×11. This inversion of pixel values is rooted in the observation that for low illumination the rain streak or snowflakes are less visible in a dark background and more pronounced near the illuminated area while the reverse is true for well-illuminated scenes.

- Color level adjustment: This involves highlighting significant rain streaks and snowflakes by identifying pixels within the range of minimum and maximum color values. Subsequently, non-selected pixels are set to 0, and the chosen pixels are rescaled to a range of 0 to 255 pixel values. The OTSU threshold is effective as a minimum threshold for automated color level adjustment.

Fog Augmentation: For each image, visibility distance, fog color, and image darkening factor are selected based on image illumination. After selecting the various parameters the image is darkened and the heterogeneous fog generation method proposed in [40] based on the Beer-Lambert law of attenuation was used to generate the fog effect. Equation 1 is used for fog attenuation, where I is the intensity after traveling distance d in fog particle, O is the opacity, and I_{al} is the fog color.

$$I_{fog} = I + O * I_{al} \tag{1}$$

where,

$$\begin{aligned} I &= I_o \exp(-\beta d) \\ O &= 1 - \exp(-\beta d) \\ \beta &= \frac{3.912}{V} m^{-1} \end{aligned} \tag{2}$$

I_o is intensity value of clear image, β is the extinction coefficient, V is the visibility in meters, and d is the distance taken from depth map D.

Rain Augmentation: For rain augmentation, we estimate the visibility distance, fog color, and image darkening factor based on image illumination and sequentially apply the darkening effect and fog attenuation with visibility greater than 500m. Subsequently, we generate a rainstreak layer (l) of the same size through a sequence of four steps. First, we create a 2D array with Gaussian noise, apply the motion blur, scale and crop the array to the original size if needed, and lastly, apply the color level adjustment to keep prominent rain streaks. By adjusting the parameters in each step, we can modify the rain streak size giving the effect of far way and near rain streaks. The resulting rain layer (l) is then blended into the image using Equation 3.

$$I_{blend} = I_o * (1 - \alpha) + l * \alpha \tag{3}$$

Snow Augmentation: Similar to the process of rain augmentation, we applied darkening and fog effects based on illumination. The procedure for generating the snowflakes layer (l) involves creating a 2D array of Gaussian noise, followed by zooming and cropping the image, applying motion blur, using color level adjustment with an OTSU threshold to enhance prominent snowflakes, and finally applying the crystallization effect. Like the rain effect generation process, adjusting parameters at each step allows us to control the size of snowflakes, resulting in effects ranging from distant to close snowfall. As with the rain effect, we employed alpha blending (Equation 3) to seamlessly integrate the snow layer into the image.

3.1.2 GAN-based Noising

In addition to the analytical weather noise generation, we employed a GAN-based approach using CycleGAN [42] to learn the mapping between clear-foggy, clear-rainy, and clear-snowy weather. The generator architecture, similar to [15], includes two downsampling blocks, nine ResNet blocks, and two upsampling blocks. The discriminator resembled the PatchGAN architecture with three hidden layers of ConvNet. To ensure context similarity of the driving dataset, we utilized adverse weather images from the Boreas Dataset [2] for training. Images were resized to a width of 512px while maintaining the aspect ratio, then randomly cropped to 224×224 dimensions. Models were trained for 100 epochs using the Adam optimizer with a batch size of 2, a learning rate of 0.0002, and $\beta = 0.5, 0.999$. The trained models generated weather effects on the same clear images used for training, keeping the original size.

3.1.3 Style-transfer Noising

We explored the application of neural style transfer, as introduced in [9], to generate weather-related noise within clear images. The algorithm takes an input image, a content image, and a style image and iteratively optimizes the input image to match the content features and the style features simultaneously. The process modifies the input image while minimizing the difference between its content and the content image's features (content loss), as well as its style and the style image's features (style loss). While the original method employs the VGG19 model, pre-trained on the ImageNet dataset for artistic style transfer, we adapted it for weather-style transfer by fine-tuning the VGG19 model as an image-based weather classifier.

3.2. Adverse Weather Denoising

We assessed various analytical and DL-based denoising methods for object detection enhancement in adverse weather. The amount of research in this field presents challenges in testing and incorporating all methods. Hence, we selected analytical and DL models (convolution and transformer-based GANs) whose implementations are available along with pre-trained weights and their frequent use for comparison in literature.

Dehazing: Extensive efforts have been directed toward enhancing foggy image restoration. The selected analytical techniques are *image_haze_removal* [43], *zero_restore* [16], and *RADE* [23]. [43] uses color attenuation prior to estimating the depth map from hazy images, which is used to get the transmission map. This transmission map aids in restoring scene radiance through atmospheric scattering modeling. The [16] method optimizes a zero-shot network to deduce parameters in Koschmieder's model, which characterizes image degradation due to light scattering in real-world scenarios. [23] algorithm splits the image into three regions: grayish sky, non-white objects, and pure white objects; then processes the non-white and non-grayish portions with luminance-inverted multi-scale Retinex accompanied by color restoration (MSRCR) and region-ratio-based adaptive Gamma correction. These processed areas are subsequently reassembled using mean-filtered region masks.

Figure 2. Result of dehazing algorithms evaluated in this work.

We selected *multi-scale-cnn-dehazing* [29], *cycle-dehaze* [7] and *FFA-Net* [27] for DL approches. *FFA-Net* [27] constitutes an end-to-end feature fusion attention network, comprising novel components like a feature attention module, a basic block structure consisting of local residual learning and feature attention, and an attention-based different levels of feature fusion structure. *multi-scale-cnn-dehazing* [29] proposes a multi-scale DL model for dehazing by learning the mapping between hazy images and their corresponding transmission maps. While *cycle-dehaze* [7] is a CyclicGAN network trained on an unpaired set of clean and hazy images.

Deraining: Deraining is one of the most researched adverse weather denoising cases in literature for analytical and DL modeling. In this work, we evaluated *UGSM* [6] and *LPDerain* [22] analytical methods, and deep learning methods like *JORDER-E* [36], *SPANet* [35], *sync2real* [37], *IPT* [3], and *DRT* [24] developed for deraining.

[22] is a decomposition-based method that uses Gaussian Model Mixture (GMM) based priors for the background and rain streak layers. An additional residue recovery step to separate the background residues is used to improve the decomposition quality. [6] formulates a simple but efficient unidirectional global sparse model UGSM for single-image rain removal. [36] introduces contextualized deep networks that handle deraining by detecting rain locations in the image, removing the rain, and finally reconstructing the image with local details without rain. [35] and [36] are DL methods that extract features, identify rain, and construct a clean output. [35] uses a novel Spatial Attentive Network while [36] uses contextualized deep networks.

[37] builds a Gaussian-Process based UNet with a semi-supervised learning framework to train a model with real-world data instead of just synthetic data. While [3] and [24] are transformer-based deraining architectures where [3] is a multi-head and multi-tail network trained on ImageNet and then fine-tuned to solve derain challenges and [24] proposes a vision transformer with a recursive local window-based self-attention structure with residual connections.

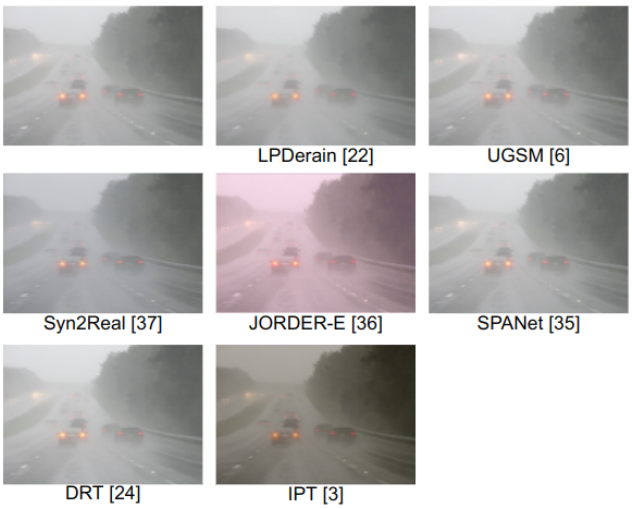

Figure 3. Denoised images using image deraining methods evaluated in this work.

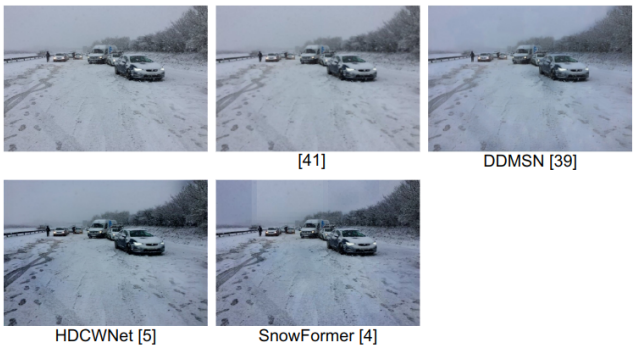

Figure 4. Denoised images using the desnowing algorithms evaluated in this work.

Desnowing: Desnowing is challenging due to the complex snow's characteristics, such as opaqueness, different shapes and sizes, uneven densities, and irregularity. We evaluated one classical method [41], and three DL-based methods *DDMSCN* [39], *HDCWNet* [5], and *SnowFormer* [4] for desnowing.

[41] creates a filter based on the area of the image that does not include snow and guides the desnowing process to create an output. [39] incorporates the semantic and geometric maps as input and learns the semantic-aware and geometry-aware representation to remove snow. [5] proposes a method to find deletable snowflakes from images based on the novel feature called the contradict channel and

clean the image. [4] is a visual transformer with a multi-head cross-attention mechanism to perform local-to-global context interaction between scale-aware snow queries and then desnowing the image using this information.

4. Dataset and Evaluation

4.1. Object Detection Dataset

Train Dataset: The Berkeley Deep Drive (BDD100K) dataset was used for training the object detection models. This comprehensive dataset offers various environmental scenarios and detailed annotations, enabling robust evaluation of adverse weather object detection strategies. It comprises 100k annotated images covering different weather conditions (clear, overcast, cloudy, rain, fog, snow, and unknown) at different times of day (daytime, night, and dusk/dawn) and split into training (70K), validation (10K) and testing (20K) subsets. The dataset has four broad object categories: vehicles (car, truck, bus, and train), humans (pedestrians and riders), bikes (bicycle and motorcycle), and miscellaneous (traffic light and signs), each image has 2D bounding box annotations and weather information.

We divided training and validation images into distinct subsets based on weather types. Two major object categories were used for training: vehicle (car, truck, bus, and train) and person (pedestrians and riders). Randomly 1500 clear weather images were selected, and various weather augmentation techniques outlined in Section 3 were applied. Approximately 1000 images were selected after manually rechecking the labels for each weather and weather augmentation approach. This resulted in five distinct training sets:

1. IMAGESET1: clear, overcast, cloudy, and unknown
2. IMAGESET2: IMAGESET1 + real-world adverse weather (fog, rain, and snow)
3. IMAGESET3: IMAGESET1 + analytical weather augmentation
4. IMAGESET4: IMAGESET1 + GAN weather augmentation
5. IMAGESET5: IMAGESET1 + style-transfer weather augmentation

Test Dataset: DAWN dataset [19] is used to assess the performance of object detection models trained using various weather augmentation techniques and adverse weather denoising methods. It encompasses severe real weather conditions like rain, fog, snow, and sandstorms. Furthermore, to assess the strategies on clear weather scenarios, the dataset from Udacity was utilized[2].

[2] https://github.com/udacity/self-driving-car/tree/master/annotations

4.2. Object Detection Model

We employed a state-of-the-art object detection model, YOLOv5 [14], from Ultralytics. This model performs a single forward pass through the neural network to simultaneously predict bounding boxes and class probabilities for objects in an image. YOLOv5 features an adaptive anchor box mechanism, enabling precise detection across various object sizes, ranging from small to large bounding boxes. It is a lightweight, faster, and memory-efficient architecture, with a codebase that can be modified effectively for model training. Our study used the pre-trained YOLOv5-l model as the initial starting point.

4.3. Training Details

Image Augmentations: Image augmentations like geometric, color, noise, and occlusion augmentations are used to enhance the model's ability to generalize and improve its robustness. In our work, three sets of augmentations were designed to facilitate the object detection model training online.

1. IMAGEAUG1: geometric augmentations (translate, scaling, and left-right flipping) and mosaic augmentation.
2. IMAGEAUG2: IMAGEAUG1 + color augmentations (HSV jittering, greying, CLACHE), noise augmentation (median blur), and occlusion augmentation (mixup).
3. IMAGEAUG3 (for mimicking weather noise): IMAGEAUG1 + color augmentations (HSV jittering, RGB Jittering, randomly adjusting brightness and contrast), noise augmentation (defocus, motion blur, Gaussian noise, pixel dropout, and image compression) and occlusion augmentation (mixup).

IMAGEAUG3 is used only with clear weather images (IMAGESET1) to study the effect of basic augmentation like motion blurring, defocus, pixel dropout, and RGB jittering on object detection in adverse weather conditions. Each image augmentation is applied with a probability of 0.01, except for geometric and mosaic augmentations.

Training Pipeline: We utilized the YOLOv5 training pipeline with an image size of 640px×640px and a batch size of 60 images. Each model is trained for 50 epochs using a stochastic gradient descent optimizer. The optimizer has a learning rate of 0.01 with cosine learning rate decay till 0.001, a momentum of 0.937, and a weight decay of 0.0005. The training loss functions encompassed box, class, and object loss functions. The training was conducted on an NVIDIA Tesla A100 graphics card, and the system comprised an Intel Xeon Gold CPU @ 2GHz with 64GB RAM.

Table 1. Object detection results for combinations of training sets and image augmentations on clear, fog, rain, and snow test images from Udacity and DAWN datasets. Training sets include IMAGESET1: clear weather, IMAGESET2: all weather, IMAGESET3: IMAGESET1+analytical weather augmentation, IMAGESET4: IMAGESET1+GAN based weather augmentation, and IMAGESET5: IMAGESET3+style-transfer weather augmentation. IMAGEAUG1: geometric augmentation and mosaic, IMAGEAUG2: IMAGEAUG1+(color augmentation, blurring, and MixUP), IMAGEAUG3: IMAGEAUG1+(color augmentation, noise augmentation, motion blur, and MixUP).

		Clear			Fog			Rain			Snow		
		mAP	P	R	mAP	P	R	mAP	P	R	mAP	P	R
IMAGESET1	IMAGEAUG1	0.474	0.834	0.734	0.506	0.855	0.704	0.478	0.763	0.735	0.472	0.848	0.690
	IMAGEAUG2	0.482	0.830	0.741	0.519	0.832	**0.788**	0.506	0.781	0.758	0.487	0.800	0.752
	IMAGEAUG3	0.480	**0.841**	0.730	0.515	0.838	0.775	0.481	**0.848**	0.706	0.489	0.781	0.750
IMAGESET2	IMAGEAUG1	0.482	**0.841**	0.730	0.506	0.838	0.713	0.469	0.824	0.713	0.505	0.809	0.771
	IMAGEAUG2	**0.484**	0.832	**0.753**	**0.523**	0.813	0.770	0.474	0.750	0.781	**0.509**	0.834	**0.776**
IMAGESET3	IMAGEAUG1	0.476	0.824	0.747	0.502	0.785	0.775	0.452	0.839	0.705	0.493	0.798	0.765
	IMAGEAUG2	0.483	0.820	0.748	0.515	0.847	0.755	0.507	0.768	0.790	0.496	0.821	0.749
IMAGESET4	IMAGEAUG1	0.478	0.825	0.738	0.497	0.836	0.758	0.476	0.769	0.714	0.484	0.845	0.744
	IMAGEAUG2	0.481	0.820	0.743	0.508	**0.858**	0.762	0.495	0.830	0.768	0.488	0.808	0.757
IMAGESET5	IMAGEAUG1	0.476	0.830	0.731	0.517	0.809	0.757	0.488	0.752	0.830	0.492	0.829	0.731
	IMAGEAUG2	0.480	0.814	0.744	0.521	0.814	0.784	**0.512**	0.737	**0.833**	0.498	**0.861**	0.730

Eleven YOLOv5-l models were trained, incorporating different combinations of training sets and image augmentation combinations.

4.4. Results

Table 1 presents the performance of YOLOv5-l model trained with different combinations of training sets (IMAGESET1 to IMAGESET5) and image augmentations (IMAGEAUG1 to IMAGEAUG3). We reported mean average precision (mAP), precision, and recall for object detection in adverse weather conditions (clear, fog, rain, and snow). The mAP represents the overall detection performance by taking into account both localization and recognition accuracy. Precision and recall offer insights into the accuracy of positive predictions and the proportion of actual positive instances correctly identified by the model. The best-performing combination is highlighted in each column.

The best mAP was observed with the model trained using real all-weather conditions (IMAGESET2), while the best precision emerged from the clear weather image set (IMAGESET1) overall. Moreover, when comparing synthetic weather augmentation, the style-transfer based train set (IMAGESET5) exhibited higher mAP in contrast to other synthetic weather augmentation methods.

We further evaluated several adverse weather denoising methods by applying denoising on test images and performing object detection on denoised images using "base" model trained with IMAGESET1+IMAGEAUG1. The mAP, precision, and recall for object detection on denoised images are reported in Table 2, Table 3, and Table 4 for dehazing, deraining, and desnowing algorithm respectively.

Across all three adverse weather scenarios, most denoising methods led to a decline in object detection performance compared to the original noisy images, with slight improvements noted in some cases. In the context of dehazing, FFA-Net [27] exhibited better performance. For deraining, IPT [3] demonstrated superior precision, while mAP was better on the original images. Similarly, regarding desnowing methods, precision proved better on noisy images, while mAP saw improvement on denoised images using the analytical approach [41]. Overall, object detection yielded better results on the original noisy images than on denoised ones.

Table 2. Result of object detection using "base" model (IMAGESET1+IMAGEAUG1) on dehazed images.

	mAP	P	R
hazy	0.506	0.855	0.704
haze-removal [43]	0.492	0.835	0.739
zero-restore [16]	0.273	0.681	0.556
RADE [23]	0.477	0.847	0.741
Cycle-Dehaze [7]	0.441	0.798	0.667
Multiscale-Dehazing [29]	0.491	0.797	**0.764**
FFA-Net [27]	**0.508**	**0.865**	0.715

5. Discussion

This study comprehensively assessed the performance of object detection models trained with various synthetic weather augmentation images in real adverse weather conditions. Additionally, we evaluated several adverse weather denoising methods by doing object detection on denoised images and comparing the results with object detection on noisy images.

Table 3. Result of object detection using "base" model (IMAGESET1+IMAGEAUG1) on derained images.

	mAP	P	R
rainy	**0.478**	0.760	0.735
LPDerain [22]	0.456	0.869	0.631
JORDER-E [36]	0.459	0.740	0.713
DRT [24]	0.466	0.809	0.701
IPT [3]	0.441	**0.871**	0.608
UGSM [6]	0.468	0.840	0.749
SPANet [35]	0.473	0.763	**0.801**
Syn2Real [37]	0.432	0.828	0.640

Table 4. Result of object detection using "base" model (IMAGESET1+IMAGEAUG1) on desnowed images.

	mAP	P	R
snowy	0.472	**0.840**	0.690
snow_removal [41]	**0.477**	0.791	**0.712**
DDMSN [39]	0.459	0.823	0.657
hdcwnet [5]	0.457	0.814	0.666
snowformer [4]	0.450	0.838	0.632

For training a robust all-weather object detection model using only clear weather images, we explored the impact of basic augmentations. While these augmentations (IMAGEAUG1, IMAGEAUG2, and IMAGEAUG3) did not significantly enhance mAP results compared to synthetic weather augmentation, they did exhibit improved precision. A reason for this enhanced precision could be that the model was trained using clear images, making the object features more distinguishable compared to the distorted features present in synthetic weather images. Among the basic image augmentations, IMAGEAUG2 yielded the best training outcomes across different training sets.

We also evaluated synthetic weather augmentations achieved through analytical methods (IMAGESET3), GAN networks (IMAGESET4), and style-transfer networks (IMAGESET5) in the context of all-weather object detection. The use of style-transfer-based weather augmentations led to enhanced mAP performance, while GAN-based augmentations yielded poorer results than training solely with the clear weather dataset (IMAGESET1). The GAN-based approach introduced significant alterations and additional information to the images, negatively impacting the results while generating complex weather noise. On the other hand, style-transfer networks add slight noise without heavily modifying the images compared to other augmentation methods, which might lead to better results. Analytical adverse weather augmentation improved object detection results slightly more than training with only clear weather images. This shows the effectiveness of weather augmentation in training a robust all-weather object detection model. It could also be used to generate a more balanced dataset with all weather conditions.

Reviewing the outcomes in Table 2, Table 3, and Table 4, it becomes apparent that most existing adverse weather denoising methods are not sufficiently compatible with object detection models. Most denoising methods resulted in the worst performance of object detection models (low mAP) on denoised images compared to the original noisy images. The reason for poor performance could be the inability of the denoising methods to add information that may aid in object detection. Instead, the denoising methods could add more noise that degrades the object detection performance, especially for the DL-based denoising algorithm. While the methods assessed in this study improved image quality by enhancing opaque object features and eliminating weather effects, this enhancement didn't consistently translate to improvements in downstream computer vision tasks. This highlights the need for alternative approaches, such as the end-to-end training of denoising and object detection models as proposed in [1] [31].

6. Conclusion

In conclusion, this study delved into enhancing object detection accuracy in adverse weather conditions. The methodologies employed are synthetic weather augmentation strategies, encompassing physics-based, GAN-based, and style-transfer approaches. The comprehensive evaluation showed that real-world all-weather conditions resulted in the best overall detection performance. At the same time, synthetic weather augmentation demonstrated its potential, with the style-transfer network emerging as particularly impactful. Moreover, the exploration of adverse weather denoising methods cast light on the intricate trade-offs between noise reduction and detection precision, underlining the inherent complexity of this task. Despite the challenges of adverse weather, the findings underscored the potential for continued advancements in object detection and denoising techniques, positioning this research as a stepping stone towards more robust and reliable computer vision systems in the face of diverse weather conditions.

Acknowledgment This work has received funding from the European Union's Horizon 2020 research and innovation287 programme under the Marie Skłodowska-Curie grant agreement No 858101.

The computations were enabled by resources provided by the National Academic Infrastructure for Supercomputing in Sweden (NAISS) at C3SE partially funded by the Swedish Research Council through grant agreement no. 2022-06725 for project SNIC 2022/5-535.

References

[1] Emmanuel Owusu Appiah and Solomon Mensah. Object detection in adverse weather condition for autonomous vehicles. *Multimedia Tools and Applications*, pages 1–27, 2023. 2, 8

[2] Keenan Burnett, David J Yoon, Yuchen Wu, Andrew Z Li, Haowei Zhang, Shichen Lu, Jingxing Qian, Wei-Kang Tseng, Andrew Lambert, Keith YK Leung, Angela P Schoellig, and Timothy D Barfoot. Boreas: A multi-season autonomous driving dataset. *The International Journal of Robotics Research*, 42(1-2):33–42, 2023. 1, 4

[3] Hanting Chen, Yunhe Wang, Tianyu Guo, Chang Xu, Yiping Deng, Zhenhua Liu, Siwei Ma, Chunjing Xu, Chao Xu, and Wen Gao. Pre-trained image processing transformer. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12299–12310, 2021. 5, 7, 8

[4] Sixiang Chen, Tian Ye, Yun Liu, Erkang Chen, Jun Shi, and Jingchun Zhou. Snowformer: Scale-aware transformer via context interaction for single image desnowing. *arXiv preprint arXiv:2208.09703*, 2022. 5, 6, 8

[5] Wei-Ting Chen, Hao-Yu Fang, Cheng-Lin Hsieh, Cheng-Che Tsai, I Chen, Jian-Jiun Ding, Sy-Yen Kuo, et al. All snow removed: Single image desnowing algorithm using hierarchical dual-tree complex wavelet representation and contradict channel loss. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 4196–4205, 2021. 5, 8

[6] Liang-Jian Deng, Ting-Zhu Huang, Xi-Le Zhao, and Tai-Xiang Jiang. A directional global sparse model for single image rain removal. *Applied Mathematical Modelling*, 59:662–679, 2018. 5, 8

[7] Deniz Engin, Anil Genç, and Hazim Kemal Ekenel. Cycle-dehaze: Enhanced cyclegan for single image dehazing. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pages 825–833, 2018. 5, 7

[8] Kshitiz Garg and Shree K Nayar. Photorealistic rendering of rain streaks. *ACM Transactions on Graphics (TOG)*, 25(3):996–1002, 2006. 1

[9] Leon A Gatys, Alexander S Ecker, and Matthias Bethge. A neural algorithm of artistic style. *arXiv preprint arXiv:1508.06576*, 2015. 4

[10] Jie Gui, Xiaofeng Cong, Yuan Cao, Wenqi Ren, Jun Zhang, Jing Zhang, Jiuxin Cao, and Dacheng Tao. A comprehensive survey and taxonomy on image dehazing based on deep learning. *arXiv e-prints*, pages arXiv–2106, 2021. 1

[11] Xiaojie Guo, Yu Li, and Haibin Ling. Lime: Low-light image enhancement via illumination map estimation. *IEEE Transactions on image processing*, 26(2):982–993, 2016. 3

[12] Xiaowei Hu, Chi-Wing Fu, Lei Zhu, and Pheng-Ann Heng. Depth-attentional features for single-image rain removal. In *Proceedings of the IEEE/CVF Conference on computer vision and pattern recognition*, pages 8022–8031, 2019. 2

[13] Xun Huang, Ming-Yu Liu, Serge Belongie, and Jan Kautz. Multimodal unsupervised image-to-image translation. In *ECCV*, 2018. 1, 2

[14] Glenn Jocher. YOLOv5 by Ultralytics, May 2020. 2, 6

[15] Justin Johnson, Alexandre Alahi, and Li Fei-Fei. Perceptual losses for real-time style transfer and super-resolution. In *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part II 14*, pages 694–711. Springer, 2016. 4

[16] Aupendu Kar, Sobhan Kanti Dhara, Debashis Sen, and Prabir Kumar Biswas. Zero-shot single image restoration through controlled perturbation of koschmieder's model. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 16205–16215, 2021. 4, 7

[17] Sotiris Karavarsamis, Ioanna Gkika, Vasileios Gkitsas, Konstantinos Konstantoudakis, and Dimitrios Zarpalas. A survey of deep learning-based image restoration methods for enhancing situational awareness at disaster sites: The cases of rain, snow and haze. *Sensors*, 22(13):4707, Jun 2022. 1

[18] Jaskirat Kaur and Williamjeet Singh. Tools, techniques, datasets and application areas for object detection in an image: a review. *Multimedia Tools and Applications*, pages 1–55, 2022. 1

[19] Mourad KENK. Dawn: Vehicle detection in adverse weather nature dataset, 2020. 1, 2, 6

[20] Siyuan Li, Iago Breno Araujo, Wenqi Ren, Zhangyang Wang, Eric K Tokuda, Roberto Hirata Junior, Roberto Cesar-Junior, Jiawan Zhang, Xiaojie Guo, and Xiaochun Cao. Single image deraining: A comprehensive benchmark analysis. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3838–3847, 2019. 1, 2

[21] Xuelong Li, Kai Kou, and Bin Zhao. Weather gan: Multi-domain weather translation using generative adversarial networks. *arXiv preprint arXiv:2103.05422*, 2021. 1, 2

[22] Yu Li, Robby T. Tan, Xiaojie Guo, Jiangbo Lu, and Michael S. Brown. Single image rain streak decomposition using layer priors. *IEEE Transactions on Image Processing*, 26(8):3874–3885, 2017. 5, 8

[23] Zhan Li, Xiaopeng Zheng, Bir Bhanu, Shun Long, Qingfeng Zhang, and Zhenghao Huang. Fast region-adaptive defogging and enhancement for outdoor images containing sky. In *2020 25th International Conference on Pattern Recognition (ICPR)*, pages 8267–8274. IEEE, 2021. 4, 7

[24] Yuanchu Liang, Saeed Anwar, and Yang Liu. Drt: A lightweight single image deraining recursive transformer. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 589–598, 2022. 5, 8

[25] Wenyu Liu, Gaofeng Ren, Runsheng Yu, Shi Guo, Jianke Zhu, and Lei Zhang. Image-adaptive yolo for object detection in adverse weather conditions. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, pages 1792–1800, 2022. 2

[26] Isaac Ogunrinde and Shonda Bernadin. A review of the impacts of defogging on deep learning-based object detectors in self-driving cars. *SoutheastCon 2021*, pages 01–08, 2021. 2

[27] Xu Qin, Zhilin Wang, Yuanchao Bai, Xiaodong Xie, and Huizhu Jia. Ffa-net: Feature fusion attention network for single image dehazing, 2019. 5, 7

[28] René Ranftl, Katrin Lasinger, David Hafner, Konrad Schindler, and Vladlen Koltun. Towards robust monocular depth estimation: Mixing datasets for zero-shot cross-dataset transfer. *IEEE transactions on pattern analysis and machine intelligence*, 44(3):1623–1637, 2020. 3

[29] Wenqi Ren, Si Liu, Hua Zhang, Jinshan Pan, Xiaochun Cao, and Ming-Hsuan Yang. Single image dehazing via multi-scale convolutional neural networks. In *European Conference on Computer Vision*, 2016. 5, 7

[30] Christos Sakaridis, Dengxin Dai, and Luc Van Gool. Semantic foggy scene understanding with synthetic data. *International Journal of Computer Vision*, 126:973–992, 2018. 2

[31] Prithwish Sen, Anindita Das, and Nilkanta Sahu. Object detection in foggy weather conditions. In *International Conference on Intelligent Computing & Optimization*, pages 728–737. Springer, 2021. 2, 8

[32] Dilbag Singh and Vijay Kumar. A comprehensive review of computational dehazing techniques. *Archives of Computational Methods in Engineering*, 26(5):1395–1413, 2019. 1

[33] Maxime Tremblay, Shirsendu Sukanta Halder, Raoul De Charette, and Jean-François Lalonde. Rain rendering for evaluating and improving robustness to bad weather. *International Journal of Computer Vision*, 129:341–360, 2021. 1, 2

[34] Alexander Von Bernuth, Georg Volk, and Oliver Bringmann. Simulating photo-realistic snow and fog on existing images for enhanced cnn training and evaluation. In *2019 IEEE Intelligent Transportation Systems Conference (ITSC)*, pages 41–46. IEEE, 2019. 2

[35] Tianyu Wang, Xin Yang, Ke Xu, Shaozhe Chen, Qiang Zhang, and Rynson W.H. Lau. Spatial attentive single-image deraining with a high quality real rain dataset. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 12262–12271, 2019. 5, 8

[36] Wenhan Yang, Robby T. Tan, Jiashi Feng, Zongming Guo, Shuicheng Yan, and Jiaying Liu. Joint rain detection and removal from a single image with contextualized deep networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 42(6):1377–1393, 2020. 5, 8

[37] Rajeev Yasarla, Vishwanath A. Sindagi, and Vishal M. Patel. Syn2real transfer learning for image deraining using gaussian processes. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2723–2733, 2020. 5, 8

[38] Fisher Yu, Haofeng Chen, Xin Wang, Wenqi Xian, Yingying Chen, Fangchen Liu, Vashisht Madhavan, and Trevor Darrell. Bdd100k: A diverse driving dataset for heterogeneous multitask learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 2636–2645, 2020. 1, 2

[39] Kaihao Zhang, Rongqing Li, Yanjiang Yu, Wenhan Luo, and Changsheng Li. Deep dense multi-scale network for snow removal using semantic and geometric priors. *IEEE Transactions on Image Processing*, 2021. 5, 8

[40] Ning Zhang, Lin Zhang, and Zaixi Cheng. Towards simulating foggy and hazy images and evaluating their authenticity. In *Neural Information Processing: 24th International Conference, ICONIP 2017, Guangzhou, China, November 14-18, 2017, Proceedings, Part III 24*, pages 405–415. Springer, 2017. 1, 2, 3

[41] Xianhui Zheng, Yinghao Liao, Wei Guo, Xueyang Fu, and Xinghao Ding. Single-image-based rain and snow removal using multi-guided filter. In Minho Lee, Akira Hirose, Zeng-Guang Hou, and Rhee Man Kil, editors, *Neural Information Processing*, pages 258–265, Berlin, Heidelberg, 2013. Springer Berlin Heidelberg. 5, 7, 8

[42] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Computer Vision (ICCV), 2017 IEEE International Conference on*, 2017. 4

[43] 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. 4, 7