WARLearn: Weather-Adaptive Representation Learning

Shubham Agarwal   Raz Birman   Ofer Hadar

Ben-Gurion University of the Negev, Israel

agarwals@post.bgu.ac.il, raz.birman@gmail.com, hadar@bgu.ac.il

Abstract

This paper introduces WARLearn, a novel framework designed for adaptive representation learning in challenging and adversarial weather conditions. Leveraging the in-variance principal used in Barlow Twins, we demonstrate the capability to port the existing models initially trained on clear weather data to effectively handle adverse weather conditions. With minimal additional training, our method exhibits remarkable performance gains in scenarios characterized by fog and low-light conditions. This adaptive framework extends its applicability beyond adverse weather settings, offering a versatile solution for domains exhibiting variations in data distributions. Furthermore, WARLearn is invaluable in scenarios where data distributions undergo significant shifts over time, enabling models to remain updated and accurate. Our experimental findings reveal a remarkable performance, with a mean average precision (mAP) of 52.6% on unseen real-world foggy dataset (RTTS). Similarly, in low light conditions, our framework achieves a mAP of 55.7% on unseen real-world low light dataset (ExDark). Notably, WARLearn surpasses the performance of state-of-the-art frameworks including FeatEnHancer, Image Adaptive YOLO, DENet, C2PNet, PairLIE and ZeroDCE, by a substantial margin in adverse weather, improving the baseline performance in both foggy and low light conditions. The WARLearn code is available at https://github.com/ShubhamAgarwal12/WARLearn

1. Introduction

One of the main limitations of computer vision based algorithms in the real-world setting is their inability to adapt to changing weather conditions. Visibility changes due to extreme weather and lighting conditions pose a significant challenge in tasks like object detection. This challenge is particularly relevant in scenarios such as autonomous cars, where inaccuracies in object recognition might lead to loss of life and property. Lately, there have been several advancements in this field. One approach that has received a lot of attention involves cleaning of the image to reduce the distortion caused by extreme weather. This includes a preprocessing step that is done using traditional image processing techniques [13, 18, 34, 40] or using deep learning based approaches [8, 10, 12, 45]. Hashmi *et al.* [12] try to learn the enhancement module weights based on the downstream task loss. Some recent methods also include additional preprocessing layers in the detection model itself [23, 29] that learn to remove the weather specific degradation. These additional steps increase the computational complexity and also the latency. Some other methods try to jointly learn image enhancement and object recognition [4, 21, 37] which makes it difficult to tune the model parameters and results in loss of performance. MAET [2] leverages adversarial learning to correctly detect objects in images modified by exposure transformations, while a generator tries to create images that challenge the detection network. There are also approaches that employ domain adaptation principles [16, 22, 33, 43]. Sindagi *et al.* [33] do this by separating the image style and weather specific features. The model then uses self supervised contrastive learning to learn features robust to adverse weather. DA-YOLO [43] also separates the features into image style and object specific. These are then separately regularised for adversarial training. Sindagi *et al.* [16] proposes cleaning the feature space by using prior-adversarial loss based on weather specific information. Li *et al.* [22] also adopts different alignment techniques for image and object level features.

WARLearn proposes a novel representation learning framework to adapt existing models trained on clean data to adverse weather conditions. Building on the idea of distribution shift in extreme weather, we learn feature representations that account for this shift. Figure 1 provides a flowchart explaining the WARLearn framework. In WARLearn, we first train the model on clean weather data. Subsequently, we extract and store the representations from the feature extraction part of the network corresponding to each clean image. Following this, we induce extreme weather conditions onto the clean data, generating a set of synthetic images under extreme weather scenarios corresponding to their clean counterparts. We then fine tune just the fea-

a) Reference Model trained on clean weather data

b) Backbone fine-tuned on adverse weather data using Barlow Loss
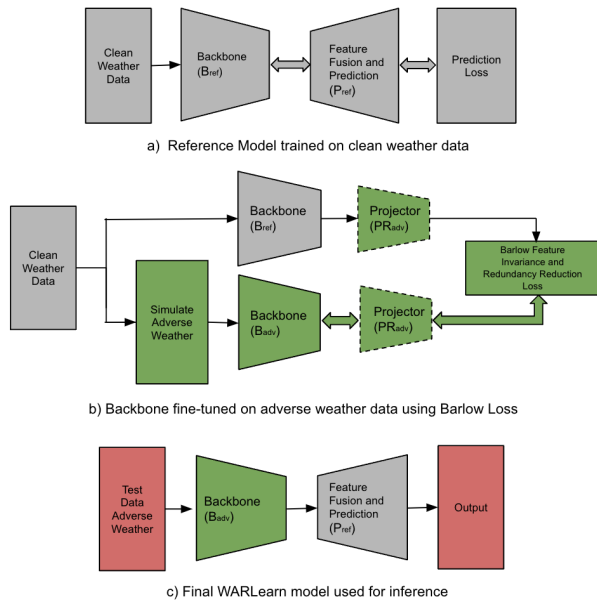
c) Final WARLearn model used for inference

Figure 1. Our proposed WARLearn framework initiates with the training of a model using clean weather data (a). Subsequently, leveraging the Barlow feature in-variance and redundancy reduction loss, we try to align the feature representations of adverse weather conditions with those obtained under clear weather conditions (b). This step enables the creation of a hybrid model, where the backbone is derived from the refined feature representations, while feature fusion and prediction components stem from the initial clean weather data (c). This hybrid model is then employed for robust predictions in adversarial weather conditions. In the figure, double-sided thick arrows indicate the flow of information between modules undergoing training in the respective stage. Dotted lines represent shared weights.

ture extractor part of the network on these synthetic extreme weather images, matching their feature representations with that of the corresponding clean images. This fine-tuning operation is facilitated through the Barlow Twins Loss [42]. The final model for inference is obtained by combining the feature extractor obtained after the fine tuning step with the prediction module obtained from the clean weather training.

In contemporary object detection research focused on challenging weather conditions, most of the latest studies [15, 23, 29] build upon the YOLOv3 [31] framework. For a fair assessment, we also adopt the YOLOv3 object detector within the WARLearn framework. We rigorously assess the performance of WARLearn on both simulated and previously unseen real-world data. Notably, with minimal fine-tuning and without any added complexity, our model either outperforms existing methods or closely approaches the state-of-the-art benchmarks, when confronted with unseen extreme weather real-world data. We provide a comprehensive comparative analysis, specifically focusing

on scenarios characterized by foggy and low-light weather conditions. We also provide experimental results using the WARLearn framework with YOLOv8 [17], demonstrating the versatility and effectiveness across different neural network architectures.

2. Related Work

Object Detection in Adverse Weather: Recently, there has been a lot of focus on vision algorithms in adverse weather conditions due to direct applicability in the real world. One of the primary research direction in this field is using the image enhancement methods or neural network based preprocessing to try and remove the effect of adverse weather [9, 13, 18, 27, 34, 38, 40, 41, 44]. Some of them solely focus on defogging [5, 32]. C2PNet [45] proposes a dehazing network combining the atmospheric scattering model and curricular contrastive regularization. Zero-DCE [10] uses non-reference loss functions to predict the pixel dynamic range adjustment for low-light image enhancement. Hahner *et al*. [11] proposed a method for simulating foggy weather on LiDAR-based 3D data. Training with this data significantly improves the 3D object detection in foggy weather. Another method that has been explored is having additional layers for preprocessing and training these together with the object recognition [23].

Valanarasu *et al*. [36] uses a transformer based architecture for image restoration. Qin *et al*. [29] proposed a preprocessing module called DENet that relies on the hypothesis that weather related information is more prominent in the low frequency component of image. This information is then used to refine the high frequency component of image that have features like edges and texture. There have also been attempts to use adversarial learning based feature calibration modules [37]. Huang *et al*. [15] introduces DSNet trained to learn tasks of visibility enhancement, object classification and object localization simultaneously. Michaelis *et al*. [25] used style transfer to mitigate the effects of adverse weather.

There are several works on using representation learning in adverse weather [14, 35, 39, 46]. Wu *et al*. [39] proposed a method that uses contrastive learning to learn feature representations that are then passed to the object recognition module. The representations improve the robustness of the model in adverse weather. Yurui *et al*. [46] try to learn weather specific and weather general features from the representations. These two kind of features are then handled differently in the model for image restoration. Sindagi *et al*. [33] proposes cleaning the feature space by using prior-adversarial loss based on weather specific information.
Barlow Twins Loss: Barlow Twins Loss [42] has been used to learn feature representations that are robust to minor distortions in the input. In this method, we first compute the correlation matrix C from the features obtained by two dis-

torted versions of the same input. Let Z^A and Z^B be the two feature vectors, mean centered along batch dimension.

$$C_{ij} \triangleq \frac{\sum_b z^A_{b,i} z^B_{b,j}}{\sqrt{\sum_b (z^A_{b,i})^2} \sqrt{\sum_b (z^B_{b,j})^2}} \tag{1}$$

where b is the index for batch samples and *i,j* are the indexes along the feature dimension. Now, the Barlow Twins loss is defined as follows:

$$L_{BT} \triangleq \sum_i (1 - C_{ii})^2 + \lambda \sum_i \sum_{j \neq i} C_{ij}^2 \tag{2}$$

where the first sum in the equation is the *invariance term* that tries to make the diagonal elements of the cross correlation matrix to 1. In doing so, it forces the network to learn distortion invariant features. The second sum, known as *redundancy reduction term* tries to make all the off diagonal elements to 0. This decorrelates different features and helps in redundancy reduction among the features. λ is a positive constant used to balance the importance of the *invariance term* and the *redundancy reduction term*.

Barlow Twins Loss has been recently used in adaptation frameworks alongside task-specific losses. Yassine *et al*. [28] use the Barlow Twins based ResNet50 feature encoder to get visual features. These features are then aligned with text features. Dripta *et al*. [30] applied it to pose estimation, involving both feature and prediction space adaptation along with regularization, where both the teacher and student models are modified during training. Similarly, Victor *et al*. [3] used it for action recognition with an action classification loss, training both source and target domain data on the same backbone. WARLearn differs from these approaches in two key aspects:
1. It uses a fixed teacher backbone as a reference, adapting only the student backbone for adverse condition.
2. It solely relies on Barlow Twins Loss for adaptation, without incorporating any task-specific losses.

3. WARLearn

The foundational idea behind our WARLearn approach centers on the premise that clear weather conditions act as a reference for object detection tasks in adverse weather conditions. When the same scene is encountered in adverse weather, performance metrics are expected to deviate from this benchmark. By training the model to discern the distribution shift from clear to adverse weather scenarios, it gains the capability to predict features for the adverse weather image as if the conditions were clear.

We start by training the model on a clear weather dataset, as illustrated in Figure 1 (a). Let's call this model M_{ref}, with its backbone denoted as B_{ref} and the predictor part as P_{ref}. Once the clear weather training is complete, we create

a) Training architecture used in Barlow Twins

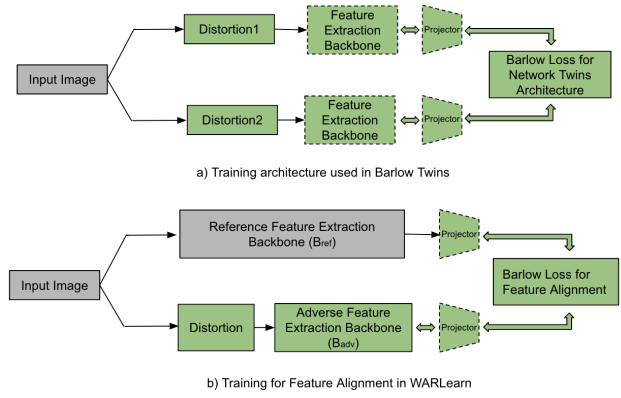b) Training for Feature Alignment in WARLearn

Figure 2. In the Barlow Twins approach, we input two distorted versions of an image into the same feature extractor. This extractor is trained to align the features of the two distorted versions, fostering similarity in their representations. WARLearn employs a foundational feature extractor initially trained on clean data as a reference. It then trains a secondary feature extractor dedicated to distorted images, utilizing the foundational extractor as a reference. The objective is to match the features extracted from distorted images with those from corresponding clean images, thereby learning to mitigate the impact of adverse weather-induced distortions. The modules that undergo training using the Barlow redundancy reduction loss are highlighted in green colour. Broken boundaries indicate weight sharing between identical modules.

a copy of B_{ref}, named B_{adv}, which is further fine-tuned for adverse weather scenario. The fine-tuning process works by first passing the clear weather image through B_{ref} and the corresponding adverse weather image through B_{adv} to obtain the reference features and adverse weather features, respectively. Both these feature sets are projected through a projection network PR_{adv} to reduce the dimensionality. Finally, we compute the Barlow loss between the projected reference features (Z^{ref}) and the projected adverse weather features (Z^{adv}). The loss is used to train the backbone B_{adv} for adverse weather. This setup is depicted in Figure 1 (b). With this feature alignment setup, Equation 1 can be rewritten as follows.

$$C_{ij} \triangleq \frac{\sum_b \boldsymbol{z^{ref}_{b,i}} z^{adv}_{b,j}}{\sqrt{\sum_b (\boldsymbol{z^{ref}_{b,i}})^2} \sqrt{\sum_b (z^{adv}_{b,j})^2}} \tag{3}$$

Notice that we now have a constant feature vector corresponding to each image in the clear weather dataset. Since the clear weather backbone B_{ref} remains fixed, the adverse weather features with varying levels of distortions are always compared against the same clear weather feature counterpart Z^{ref}. Assuming a single sample in a batch, the diagonal element in the correlation matrix C_{ii} becomes the dot product of the feature set from the clear image and the adverse weather image. The invariance term in the Barlow

Twins loss becomes as follows:

$$L_{invariance} = (1 - \boldsymbol{Z^{ref}} \cdot Z^{adv})^2 \qquad (4)$$

As we can see in Equation 4, the invariance loss term is minimized to make the dot product $Z^{ref} \cdot Z^{adv}$ close to 1, thus introducing similarity between the adverse weather image features through B_{adv} and the clear weather features through B_{ref}. This ensures that the weights of B_{adv} learn to mitigate the effect of adverse conditions and produce features as if the input image were clear and passed through B_{ref}. The redundancy reduction term in Equation 2 is used to introduce decorrelation among the feature set, which helps in generalization. Figure 2 highlights the difference between the Barlow Twins and the WARLearn framework.

The final model for inference is depicted in Figure 1(c), where the backbone B_{adv} is obtained from step Figure 1 (b) to mitigate the effect of adverse weather, and the P_{ref} module remains the same as that in Figure 1 (a). This setup has inherent modularity, and the same P_{ref} module can be utilized for multiple adversarial scenarios by plugging in the corresponding B_{adv}.

The WARLearn framework can work under two conditions. Firstly, the model needs to have separable feature extraction and prediction modules. Secondly, we need a simulation algorithm to apply the desired distortions on the clear image dataset so that there is image and feature correspondence for Barlow loss based feature alignment. With these conditions satisfied, our framework can be extended easily to other adverse scenarios.

4. Experiment Setup

We conducted experiments using the WARLearn framework on the YOLOv3 model with the Darknet-53 backbone [31]. Darknet-53 is a sequential, 53 layer deep network that features efficient feature extraction, residual connections, and multi-scale feature integration. The Darknet-53 module serves as our feature extractor.

The base YOLOv3 model with clean data was trained on the widely used Pascal VOC dataset [6,7], which comprises 20 classes: *aeroplane, bicycle, bird, boat, bottle, bus, car, cat, chair, cow, dining-table, dog, horse, motorbike, person, potted-plant, sheep, sofa, train, tv-monitor*. The dataset contains 19,561 training images and 4,952 test images. For training, we utilized a batch size of 32 and the ADAM optimizer [19] with an initial learning rate of 0.02. The experiments were set up using PyTorch, and we employed an NVIDIA GEFORCE RTX 3090, 32 GB GPU for execution. The clean data training was performed for 600 epochs.

To simulate adverse weather conditions, specifically foggy weather and low light, on the PascalVOC dataset, we followed the methodology described in [37]. Foggy images

were generated using the atmospheric scattering model developed by [1, 26], where:

$$I(x) = J(x)t(x) + A(1 - t(x)) \qquad (5)$$

Here, $I(x)$ is the foggy image, $J(x)$ is the clean image, A is the global atmospheric light, and $t(x)$ is the medium transmission map obtained using the scattering coefficient of the atmosphere β and scene depth $d(x)$:

$$t(x) = e^{-\beta}d(x) \qquad (6)$$

$$d(x) = -0.04\rho + \sqrt{\max(\text{rows}, \text{cols})} \qquad (7)$$

Here, ρ is the distance of the current pixel from the center, and rows and cols are the number of rows and columns in the image, respectively. We set A to 0.5 and

$$\beta = 0.01i + 0.05 \qquad (8)$$

We vary i from 0 to 9 to obtain 10 levels of foggy images for each clean image in the PascalVOC dataset.

Similarly, for obtaining low-light images, we used the transformation:

$$l(x) = x^{\gamma} \qquad (9)$$

Here, x is the intensity of the input pixel, and γ is randomly sampled from a uniform distribution within the range [1.5, 5]. We randomly sampled 10 different γ values to obtain 10 different low-light images corresponding to each image in the PascalVOC dataset.

In the original Barlow Twins paper [42], experiments show that the loss (Equation 2) is not very sensitive to the hyper-parameter λ. We adopt the λ value of 0.001 for all our experiments. To highlight the significance of the redundancy reduction term, we include a comparison of WARLearn with $\lambda = 0$ in the ablation studies.

During fine-tuning for adverse weather, we reduced the learning rate to 0.0002 and fine-tuned the backbone for 10 epochs. To evaluate performance on real-world data, we used the RTTS [20] and ExDark [24] datasets. It's important to note that these datasets were not used during training or fine-tuning. Table 2 lists the common classes between these datasets and the PascalVOC dataset. Note that we train all the WARLearn models with all the 20 classes in PascalVOC dataset.

Many of the SOTA methods for object detection in adverse weather either use the model trained on the training set of ExDark or sample the PascalVOC clear/simulated weather data to include samples with the common classes (Table 2) present in ExDark [24] for low-light and RTTS [20] for foggy weather. We instead train or fine-tune the the WARLearn model with all the 20 classes present in the PascalVOC dataset. For a fair performance comparison, we

Dataset	Type	Images	Classes	Instances	Purpose
PascalVOC train	Real	19561	20	40058	Clear weather training/Part of mixed
PascalVOC train foggy	Simulated	195610	20	400580	Part of foggy mixed
PascalVOC train lowlight	Simulated	195610	20	400580	Part of lowlight mixed
PascalVOC train foggy mixed	Real+Simulated	215171	20	440638	Foggy weather training
PascalVOC train lowlight mixed	Real+Simulated	215171	20	440638	Lowlight weather training
PascalVOC test	Real	4952	20	12032	Clear weather testing
PascalVOC test foggy (SimFoggy)	Simulated	49520	20	120320	Foggy weather testing
PascalVOC test lowlight (SimLL)	Simulated	49520	20	120320	Lowlight weather testing
RTTS	Real	4322	10	29599	Unseen Foggy weather testing
ExDark	Real	2563	5	6450	Unseen Lowlight weather testing

Table 1. Datasets.

Dataset	Common Classes
PascalVOC	all
RTTS	person, bicycle, car, bus, motorcycle
ExDark	person, bicycle, boat, bottle, bus, car, cat, chair, dog, motorcycle

Table 2. Common classes with the PascalVOC dataset.

Framework	Train Data	Clean	SimFoggy	RTTS
YOLOv3	PascalVOC train	76.30	53.61	46.20
YOLOv3	PascalVOC foggy mixed	66.80	71.20	49.00
C2PNet + YOLOv3	PascalVOC train	—-	66.50	47.70
IA-YOLO	PascalVOC foggy mixed	69.38	67.58	47.68
DENet	PascalVOC foggy mixed	68.40	65.60	50.50
C2PNet + WARLearn	PascalVOC foggy mixed	—-	69.30	**52.00**
WARLearn	PascalVOC foggy mixed	69.11	**75.10**	**52.60**

Table 3. Comparison of mAP@50 (%) for various frameworks on Clean (PascalVOC test), SimFoggy (PascalVOC test foggy), and real-world foggy RTTS datasets. Last two rows are experiments with our new WARLearn framework.

retrain other SOTA methods on the same dataset as WAR-Learn. The dataset details are provided in Table 1.

To prove the versatility of WARLearn framework, we also conduct similar experiments with the ultralytics YOLOv8 [17] model on the same datasets. The clean weather training was done for 100 epochs. The starting learning rate was 0.01 for the clear weather training and was subsequently reduced by a factor to 100 to 0.0001 for the adverse weather fine tuning. The fine-tuning for YOLOv8 was also done for 10 epochs. We present these results as part of the ablation studies.

5. Results

We test the WARLearn framework on adverse weather scenarios of foggy and low light conditions.

5.1. Results with Foggy Weather Datasets

Table 3 presents the results on foggy test data, showcasing the superior performance of the WARLearn framework compared to other detection frameworks on both synthetic and real-world foggy datasets.

The WARLearn framework surpasses end to end detectors like IA-YOLO [23] and DENet [29] by a huge margin. The mean Average Precision (mAP) value of 75.10% on synthetic foggy data surpasses the closest competitor by 3.90%. Similarly, the mAP value of 52.60% on the RTTS [20] dataset exceeds the nearest competitor by 2.10%. The drop in clear weather performance of WARLearn is minimal

(0.27% less) when compared to other frameworks trained on mixed foggy data. Even with SOTA pre-processing method like C2PNet [45], WARLearn performs better than the YOLOv3 model on both simulated foggy and real world RTTS datasets. This is due to the fact that the pre-processing does not remove all the foggy artifacts and the weights of WARLearn are learnt with mixed training data containing both clear and foggy weather images. The redundancy reduction term in the WARlearn also helps to learn more robust features.

Figure 3 shows the performance comparison of YOLOv3 and WARLearn on a few sample images from the RTTS dataset.

Figure 3. Comparison of detection results between YOLOv3 baseline (left column) and WARLearn (right column) on the real-world foggy RTTS dataset. WARLearn recognizes more objects in foggy conditions with higher confidence scores.

5.2. Results with Low Light Weather Datasets

The results on low-light weather datasets mirror those obtained on foggy weather datasets. Table 4 details these results for low-light weather datasets. We obtain mAP@50 of 70.90% with the WARLearn framework on the synthetic low-light PascalVOC dataset and 55.70% percent on the real-world ExDark [24] dataset. WARLearn surpasses the nearest mAP@50 by 2.70% and 2.00% on the synthetic and real world low light datasets, respectively. Here, the clear weather performance is also substantially better(3.16% more) than other generalized frameworks trained on mixed low-light dataset. Figure 4 shows the performance comparison of YOLOv3 and WARLearn on some sample images from the WARLearn dataset.

Figure 4. Comparison of detection results between YOLOv3 baseline (left column) and WARLearn (right column) on the real-world low-light ExDark dataset. WARLearn excels in recognizing more objects in low-light conditions with higher confidence scores.

In lowlight conditions also, the SOTA pre-processing methods of ZeroDCE [10] and PairLIE [8] perform better

Framework	Train Data	Clean	SimLL	ExDark
YOLOv3	PascalVOC train	76.30	59.90	49.30
YOLOv3	PascalVOC lowlight mixed	70.20	67.60	49.90
PairLIE + YOLOv3	PascalVOC train	—-	60.10	42.70
ZeroDCE + YOLOv3	PascalVOC train	—-	63.10	48.10
IA-YOLO	PascalVOC lowlight mixed	72.34	63.38	48.16
DENet	PascalVOC lowlight mixed	72.20	62.40	47.50
PairLIE + WARLearn	PascalVOC lowlight mixed	—-	63.90	47.00
ZeroDCE + WARLearn	PascalVOC lowlight mixed	—-	67.00	51.30
FeatEnhancer (YOLOv3)	PascalVOC lowlight mixed	—-	68.20	53.70
WARLearn	PascalVOC lowlight mixed	75.50	**70.90**	**55.70**

Table 4. Comparison of mAP@50 (%) for various frameworks on Clean (PascalVOC test), SimLL (PascalVOC test lowlight), and real-world low-light ExDark datasets. Last two rows are experiments with WARLearn framework.

with WARLearn than the YOLOv3. We can also observe that IA-YOLO, DENet, PairLIE and ZeroDCE+YOLOv3, do not generalize well on unseen real world foggy RTTS dataset and have a slightly lower mAP@50 compared to baseline YOLOv3. However they perform well on the synthetic PascalVOC lowlight test data that is similar to the data used for training. FeatEnHancer [12] is the closest and generalises well, still falling short of WARLearn performance. This result also indicates that along with the invariance term that aligned the adverse weather features to clear features in the WARLearn, redundancy reduction loss term is also useful as for better generalization on real world data.

In general, we note a substantial improvement in performance when utilizing the WARLearn framework compared to other state-of-the-art frameworks on adverse weather datasets. Since the WARLearn model is fine-tuned on synthetic adverse weather data, we observe a steep increase in mAP values for the synthetic adverse weather test data. This is without a significant drop in performance on clear weather dataset. Furthermore, we observe a considerable increase in mAP when applied to real-world adverse weather datasets.

6. Ablation Studies

6.1. Experiments with YOLOv8

To prove the effectiveness of the WARLearn framework across different architectures, we did the experiments with the latest YOLOv8 [17] framework from ultralytics.

Framework	Train Data	Clean	SimFoggy	RTTS
YOLOv8	PascalVOC foggy mixed	78.90	76.00	58.20
YOLOv8-WARLearn	PascalVOC foggy mixed	79.70	**76.60**	**59.80**

Table 5. Comparison of mAP@50 (%) for YOLOv8 and YOLOv8-WARLearn on Clean (PascalVOC test), SimFoggy (PascalVOC test foggy), and real-world foggy RTTS datasets.

Framework	Train Data	Clean	SimLL	ExDark
YOLOv8	PascalVOC lowlight mixed	81.30	73.60	57.70
YOLOv8-WARLearn	PascalVOC lowlight mixed	79.80	71.00	**60.00**

Table 6. Comparison of mAP@50 (%) for YOLOv8 and YOLOv8-WARLearn on Clean (PascalVOC test), SimLL (PascalVOC test lowlight), and real-world lowlight ExDark datasets.

Under both foggy conditions Table 5 and lowlight conditions Table 6, we observe significant improvements when employing the WARLearn framework with the YOLOv8 model on real world adverse weather datasets. These results demonstrate the versatility and effectiveness of the WARLearn framework, highlighting it's adaptability to various model architectures. This adaptability is not constrained by current models and can be seamlessly integrated with the latest advancements in model development.

6.2. Parameters and Inference Time

We conducted a comparative analysis of the inference time for various SOTA approaches alongside our proposed WARLearn method. For this comparison, we ran the inference time measurements on a 64-bit Windows machine equipped with an Intel Core i9-10900 CPU, 64 GB of RAM, and an NVIDIA GEFORCE RTX 3090 GPU with 32 GB of memory. The input image size is 416x416 for all the frameworks. For better understanding, we also compared the number of model parameters across these frameworks. The results of this analysis underscore are provided in Table 7.

We can see that all the adverse weather detection frameworks including IA-YOLO [23] and DENet [29] require additional parameters and inference time compared to YOLOv3. C2PNet [45], PairLIE [8], FeatEnHancer [12]

Framework	Additional Parameters	Additional Inference Time (in ms)
YOLOv3	–	–
C2PNet	7.16M	77.00
PairLIE	340K	49.03
ZeroDCE	79K	6.31
IA-YOLO	165K	7.20
FeatEnHancer	138K	6.84
DENet	45K	5.04
WARLearn	**0**	**0.00**

Table 7. Comparison of Parameters and Inference Time for different adverse weather detection frameworks. For this study, we consider YOLOv3 as a baseline and include only the additional parameteres and inference time when compared to YOLOv3.

and ZeroDCE [10] rely on pre-processing using additional parameters. WARLearn does not introduce any additional parameters and uses the same architecture as YOLOv3, ensuring that inference time remains same as that of YOLOv3. This characteristic highlights the efficiency of WARLearn in maintaining computational performance. This shows the WARLearn's ability to deliver competitive inference times with a significant boost in performance, making it a viable option for real-time object detection tasks, especially under adverse conditions.

6.3. Performance with Different Degradation Levels

We also conducted experiments to observe the variation in performance across different levels of noise under foggy weather conditions. For this experiment, we divided the SimFoggy dataset into ten parts, each representing different levels of degradation with i varying from 0 to 9 in Equation (8). This results in β ranging from 0.05 to 0.14. Each of these ten parts contains 4952 images, with a specific β value applied to the clear PascalVOC test dataset.

Fog Level β	Precision	Recall	mAP@50	F1
0.05	0.658	0.779	0.751	0.711
0.06	0.641	0.786	0.756	0.704
0.07	0.629	0.799	0.760	0.702
0.08	0.621	0.802	0.759	0.697
0.09	0.613	0.807	0.761	0.694
0.10	0.602	0.807	0.760	0.687
0.11	0.596	0.806	0.761	0.682
0.12	0.584	0.805	0.757	0.673
0.13	0.577	0.800	0.750	0.667
0.14	0.566	0.796	0.742	0.658

Table 8. Performace comparison of WARLearn at different levels of fog.

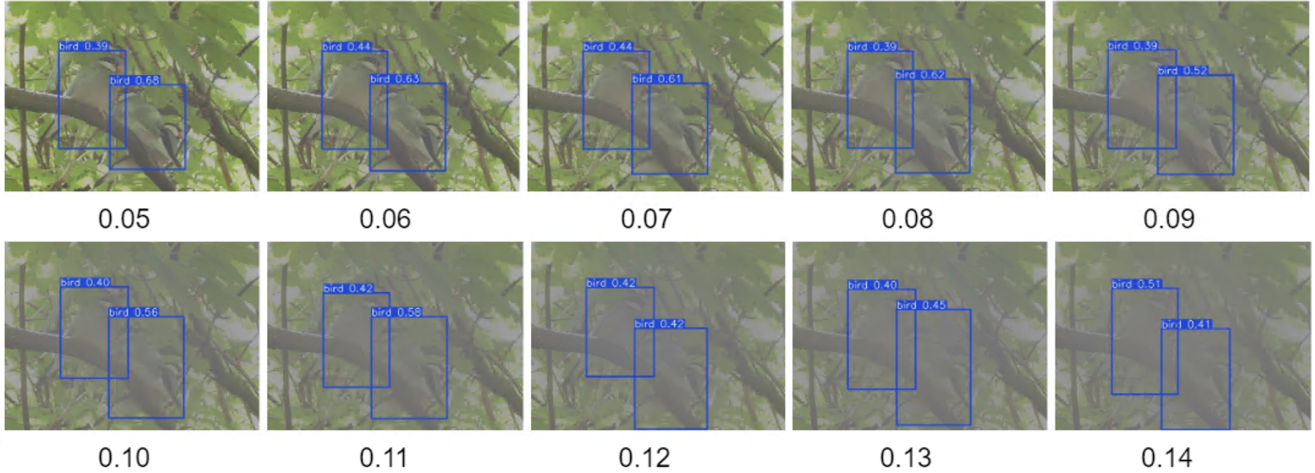As shown in Table 8, the precision and F1 score exhibit

Figure 5. Detection results with WARLearn framework on an image with different levels of foggy degradation indicated by β (Equation (8)) value below the image. These images are taken from simulated PascalVOC test foggy dataset.

a consistent decline with increasing levels of degradation, which is expected given their sensitivity to the detection threshold. However, mAP@50 remains consistent across various levels of fog. This is further illustrated in Figure 5, where the detection performance is similar regardless of the degradation level. There is a slight change in the bounding box at higher levels of noise, but the network consistently identifies two birds in the images, even when they are barely visible to the human eye. We observed comparable performance of WARLearn with other images as well. The few instances where WARLearn failed to detect objects occurred when the objects were far from the camera and became indistinguishable due to degradation. This analysis demonstrates that the WARLearn framework can learn robust features and maintain consistent mAP@50 performance with varying degrees of noise.

6.4. Significance of Redundancy Reduction Term

We conducted experiments on foggy dataset using WARLearn with λ (Equation 2) being 0 such that we only use the invariance term for adaptation. We can clearly see in Table 9, that the redundancy reduction term substantially improves the WARLearn performance in adverse weather with a minor drop in performance on clear weather dataset.

7. Conclusion

WARLearn introduces a novel and efficient framework to facilitate object detection in adverse weather conditions. The experimental results validate our approach, showcasing superior performance over the existing state-of-the-art methods on both simulated and real world adverse weather datasets. The framework is also efficient and has same inference time and parameters as baseline YOLOv3.

λ	Condition	Clean	SimFoggy	RTTS
0.000	No redundancy reduction	69.60	67.40	47.90
0.001	With redundancy reduction	69.11	75.10	52.60

Table 9. Comparison of mAP@50 (%) with and without redundancy reduction in WARLearn on Clean (PascalVOC test), SimFoggy (PascalVOC test foggy), and real-world foggy RTTS datasets. We use PascalVOC foggy mixed dataset for training these models.

In our experiments, we demonstrated the effectiveness of the WARLearn framework in handling adverse conditions such as lowlight and foggy weather. However, the versatility of the WARLearn framework extends beyond these specific scenarios. As long as we can simulate an adverse condition on clean data and have a detection model with a backbone, such as YOLO, to extract the features, WARLearn can be employed to mitigate the effects of these adverse conditions and accurately detect objects. It is also well-suited for scenarios where there is a distribution shift in data over time. Future work can explore the integration of WARLearn with other state-of-the-art detection models and it's application in a broader range of challenging real-world scenarios. By doing so, we can further validate it's effectiveness, leading to development of more resilient and adaptive computer vision systems.

8. Acknowledgment

References

[1] A. Cantor. Optics of the atmosphere–scattering by molecules and particles. *IEEE Journal of Quantum Electronics*, 14(9):698–699, 1978. 4

[2] Ziteng Cui, Guo-Jun Qi, Lin Gu, Shaodi You, Zenghui Zhang, and Tatsuya Harada. Multitask aet with orthogonal tangent regulary for dark object detection. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 2553–2562, October 2021. 1

[3] Victor G Turrisi Da Costa, Giacomo Zara, Paolo Rota, Thiago Oliveira-Santos, Nicu Sebe, Vittorio Murino, and Elisa Ricci. Unsupervised domain adaptation for video transformers in action recognition. In *2022 26th International Conference on Pattern Recognition (ICPR)*, pages 1258–1265. IEEE, 2022. 3

[4] Qiqi Ding, Peng Li, Xuefeng Yan, Ding Shi, Luming Liang, Weiming Wang, Haoran Xie, Jonathan Li, and Mingqiang Wei. Cf-yolo: Cross fusion yolo for object detection in adverse weather with a high-quality real snow dataset. *IEEE Transactions on Intelligent Transportation Systems*, 2023. 1

[5] 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 *CVPR*, pages 2154–2164, 2020. 2

[6] M. Everingham., L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman. The PASCAL Visual Object Classes Challenge 2007 (VOC2007) Results. http://www.pascal-network.org/challenges/VOC/voc2007/workshop/index.html. 4

[7] M. Everingham., L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman. The PASCAL Visual Object Classes Challenge 2012 (VOC2012) Results. http://www.pascal-network.org/challenges/VOC/voc2012/workshop/index.html. 4

[8] Zhenqi Fu, Yan Yang, Xiaotong Tu, Yue Huang, Xinghao Ding, and Kai-Kuang Ma. Learning a simple low-light image enhancer from paired low-light instances. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 22252–22261, 2023. 1, 6, 7

[9] Jeong gi Kwak, Youngsaeng Jin, Yuanming Li, Dongsik Yoon, Donghyeon Kim, , and Hanseok Ko. Adverse weather image translation with asymmetric and uncertainty-aware gan. 2021. 2

[10] Chunle Guo Guo, Chongyi Li, Jichang Guo, Chen Change Loy, Junhui Hou, Sam Kwong, and Runmin Cong. Zero-reference deep curve estimation for low-light image enhancement. In *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)*, pages 1780–1789, June 2020. 1, 2, 6, 7

[11] Martin Hahner, Christos Sakaridis, Dengxin Dai, and Luc Van Gool. Fog simulation on real lidar point clouds for 3d object detection in adverse weather. pages 15283–15292, 2021. 2

[12] Khurram Azeem Hashmi, Goutham Kallempudi, Didier Stricker, and Muhammad Zeshan Afzal. Featenhancer: Enhancing hierarchical features for object detection and beyond under low-light vision. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 6725–6735, October 2023. 1, 6, 7

[13] Mahmoud Hassaballah, Mourad A. Kenk, Khan Muhammad, and Shervin Minaee. Vehicle detection and tracking in adverse weather using a deep learning framework. In *IEEE transactions on intelligent transportation systems 22*, pages 4230–4242, 2020. 1, 2

[14] Robin Heinzler, Florian Piewak, Philipp Schindler, and Wilhelm Stork. Cnn-based lidar point cloud de-noising in adverse weather. *IEEE Robotics and Automation Letters*, 10.1109/LRA.2020.2972865:2377–3774, 2020. 2

[15] Shih-Chia Huang, Trung-Hieu Le, , and Da-Wei Jaw. Dsnet: Joint semantic learning for object detection in inclement weather conditions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43(8):2623–2633, 2020. 2

[16] Minsik Jeon, Junwon Seo, and Jihong Min. Da-raw: Domain adaptive object detection for real-world adverse weather conditions. *ArXiv*, abs/2309.08152, 2023. 1

[17] Glenn Jocher, Ayush Chaurasia, and Jing Qiu. Ultralytics YOLO., Jan. 2023. 2, 5, 7

[18] Sanket Kalwar, Dhruv Patel, Aakash Aanegola, Krishna Reddy Konda, Sourav Garg, and K. Madhava Krishna. Gdip: Gated differentiable image processing for object detection in adverse conditions. In *ICRA*, pages 7083–7089, 2023. 1, 2

[19] Diederik P. Kingma and Jimmy Lei Ba. Adam: A mthod for stochastic optimization. 2015. 4

[20] 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, 2019. 4, 5

[21] Chengyang Li, Heng Zhou, Yang Liu, Caidong Yang, Yongqiang Xie, Zhongbo Li, and Liping Zhu. Detection-friendly dehazing: Object detection in real-world hazy scenes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023. 1

[22] Jinlong Li, Runsheng Xu, Jin Ma, Qin Zou, Jiaqi Ma, and Hongkai Yu. Domain adaptive object detection for autonomous driving under foggy weather. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 612–622, 2023. 1

[23] 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*, 2022. 1, 2, 5, 7

[24] Yuen Peng Loh and Chee Seng Chan. Getting to know low-light images with the exclusively dark dataset. *Computer Vision and Image Understanding*, 178:30–42, 2019. 4, 6

[25] Claudio Michaelis, Benjamin Mitzkus, Robert Geirhos, Evgenia Rusak, Oliver Bringmann, Alexander S. Ecker, Matthias Bethge, and Wieland Brendel. Benchmarking robustness in object detection: Autonomous driving when winter is coming. *arXiv preprint arXiv:1907.07484*, 2019. 2

[26] S.G. Narasimhan and S.K. Nayar. Vision and the atmosphere. *International Journal of Computer Vision*, 14(9):233–254, 2002. 4

[27] S.G. Narasimhan and S.K. Nayer. Contrast restoration of weather degraded images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(6):713–724, 2003. 2

[28] Yassine Ouali, Adrian Bulat, Brais Martinez, and Georgios Tzimiropoulos. Black box few-shot adaptation for vision-language models, 2023. 3

[29] Qingpao Qin, Kan Chang, Mengyuan Huang, and Guiqing Li. Denet: Detection-driven enhancement network for object detection under adverse weather conditions. 2022. 1, 2, 5, 7

[30] Dripta S Raychaudhuri, Calvin-Khang Ta, Arindam Dutta, Rohit Lal, and Amit K Roy-Chowdhury. Prior-guided source-free domain adaptation for human pose estimation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 14996–15006, 2023. 3

[31] Joseph Redmon and Ali Farhadi. Yolov3: An incremental improvement. *arXiv*, 2018. 2, 4

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

[33] Vishwanath A Sindagi, Poojan Oza, Rajeev Yasarla, and Vishal M Patel. Prior-based domain adaptive object detection for hazy and rainy conditions. In *ECCV*, pages 763–780. Springer, 2020. 1, 2

[34] Robby T. Tan. Visibility in bad weather from a single image. In *CVPR*, pages 1–8, 2008. 1, 2

[35] Zhentao Tan, Yue Wu, Qiankun Liu, Qi Chu, Le Lu, Jieping Ye, and Nenghai Yu. Exploring the application of large-scale pre-trained models on adverse weather removal. *arXiv preprint arXiv:2306.09008*, 2019. 2

[36] Jeya Maria Jose Valanarasu, Rajeev Yasarla, and Vishal M. Patel. Transweather: Transformer-based restoration of images degraded by adverse weather conditions. 2022. 2

[37] Lucai Wang, Qin Hongda, Zhou Xuanyu, Lu Xiao, , and Zhang Fengting. R-yolo: A robust object detector in adverse weather. In *IEEE Transactions on Instrumentation and Measurement 72*, pages 1–11, 2022. 1, 2, 4

[38] Chen Wei, Wenjing Wang, Wenhan Yang, and Jiaying Liu. Deep retinex decomposition for low-light enhancement. *arXiv preprint arXiv:1808.04560*, 2018. 2

[39] Wei Wu, Hao Chang, Yonghua Zheng, Zhu Li, Zhiwen Chen, and Ziheng Zhang. Contrastive learning-based robust object detection under smoky conditions. In *CVPR Workshops*, pages 4295–4302, 2022. 2

[40] Jiale Yao, Fan Xiangsuo, Li Bing, and Qin Wenlin. Adverse weather target detection algorithm based on adaptive color levels and improved yolov5. In *Sensors 22*, volume 21, 2022. 1, 2

[41] Tian Ye, Sixiang Chen, Jinbin Bai, Jun Shi, Chenghao Xue, Jingxia Jiang, Junjie Yin, Erkang Chen, and Yun Liu. Adverse weather removal with codebook priors. 2023. 2

[42] Jure Zbontar, Li Jing, Ishan Misra, Yann LeCun, and Stéphane Deny. Barlow twins: Self-supervised learning via redundancy reduction. In *ICML*, 2021. 2, 4

[43] Shizhao Zhang, Hongya Tuo, Jian Hu, and Zhongliang Jing. Domain adaptive yolo for one-stage cross-domain detection. In *Asian Conference on Machine Learning*, pages 785–797. PMLR, 2021. 1

[44] Yonghua Zhang, Jiawan Zhang, and Xiaojie Guo. Kindling the darkness: A practical low-light image enhancer. In *Proceedings of the 27th ACM international conference on multimedia*, pages 1632–1640, 2019. 2

[45] Yu Zheng, Jiahui Zhan, Shengfeng He, Junyu Dong, and Yong Du. Curricular contrastive regularization for physics-aware single image dehazing. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023. 1, 2, 5, 7

[46] Yurui Zhu, Tianyu Wang, Xueyang Fu, Xuanyu Yang, Xin Guo, Jifeng Dai, Yu Qiao, and Xiaowei Hu. Represent, compare, and learn: A similarity-aware framework for class-agnostic counting. In *CVPR*, 2023. 2