



## Investigating Weather-Adaptive Dehazing Effects on YOLOv9 Object Detection in Adverse Weather

Arif Agung Saptoto<sup>1,2\*</sup>, Pulung Nurtantio Andono<sup>1</sup>, Abdul Syukur<sup>1</sup>, Affandy<sup>1</sup>, Aris Marjuni<sup>1</sup>

<sup>1</sup> Faculty of Computer Science, Universitas Dian Nuswantoro, Semarang 50131, Indonesia

<sup>2</sup> Universitas Islam Kalimantan Muhammad Arsyad Al Banjari, Banjarmasin 70114, Indonesia

Corresponding Author Email: [p41202300072@mhs.dinus.ac.id](mailto:p41202300072@mhs.dinus.ac.id)

Copyright: ©2026 The authors. This article is published by IETA and is licensed under the CC BY 4.0 license (<http://creativecommons.org/licenses/by/4.0/>).

<https://doi.org/10.18280/jesa.590204>

### ABSTRACT

**Received:** 16 December 2025

**Revised:** 14 February 2026

**Accepted:** 24 February 2026

**Available online:** 28 February 2026

#### Keywords:

*object detection, adverse weather conditions, YOLOv9, DAWN dataset, trade-off, Dark Channel Prior, Contrast Limited Adaptive Histogram Equalization, autonomous driving*

Object detection struggles in adverse weather. Rain obscures through droplets and motion blur, fog scatters atmospheric light, and sandstorms generate heavy particulate haze. Snow whitewashes scenes. Existing approaches target architectural improvements (attention mechanisms, modified losses) or pre-processing strategies (dehazing, restoration). Yet systematic investigation of how weather-specific dehazing affects the relationship between detection completeness and localization precision remains absent. We apply weather-adaptive dehazing to the Detection in Adverse Weather Nature (DAWN) dataset: Dark Channel Prior (DCP) for fog, Contrast Limited Adaptive Histogram Equalization (CLAHE)-based enhancement for rain, sand, and snow, feeding processed images into YOLOv9-C. Baseline established through stratified weather splitting achieves 69.4% mAP@0.5, exceeding prior work by 21 percentage points. Weather-adaptive dehazing reveals a fundamental trade-off between detection completeness and localization accuracy. Recall increases 12.3%, precision rises 7.1%, yet mAP drops 14.7%. The model detects more objects with fewer false positives, but localization accuracy degrades because preprocessing alters feature distributions in ways that enhance object discriminability while disrupting spatial precision. Motorcycle benefits (+3.3 AP@0.5), bicycle collapses catastrophically (-49.4 AP@0.5), car remains robust. Snow doubles detection counts (+127) alongside steep confidence loss (-7.9 points). Fog reduces detections (-69) while maintaining stable confidence. Thin structures fail under edge artifacts, dense objects gain from enhanced visibility. Findings guide practical deployment: apply dehazing for small dense objects in snow or sand conditions, avoid for thin structures or fog scenarios where classical methods introduce harmful artifacts.

## 1. INTRODUCTION

Road traffic monitoring systems depend on detecting pedestrians, motorcycles, cars, buses, and trucks. Accurate detection prevents accidents, monitors congestion, and improves operational efficiency [1]. Weather heavily influences outdoor detection, particularly on highways. Object detection algorithms perform well under clear conditions, yet degrade in adverse weather such as fog, rain, sand, or snow. Visual feature extraction becomes limited, plus the generalization suffers [2, 3], requiring detection adaptive to environmental variation [4]. Location uncertainty and perception bias emerge from decreased visual sensor quality. Weather disturbances demand robust models [5, 6]. It is worth noting that the weather heavily influences outdoor detection, particularly on highways. Rain reduces visibility through water droplets on camera lenses plus motion blur from falling rain. Fog creates atmospheric scattering that washes out contrast. Sandstorms introduce dense particulate matter blocking light transmission. Snow produces overexposure and color distortion through high surface reflectance. These degradations cause missed detections, false positives, and

localization errors that compromise system reliability.

Image restoration techniques have gained attention in computer vision through dehazing, de-raining, de-snowing, dust removal, and denoising. Visibility improvement drives most approaches. The Dark Channel Prior (DCP) remains widely used [7]. Yet detection-friendly dehazing and restoration-enhanced methods show that pre-processing affects downstream detection differently across weather types [8, 9]. Some studies report gains, others find minimal impact or even degradation.

Object detection in adverse weather still remains a major challenge. The process requires pre-processing that can (1) recognize the dominant weather disturbance, (2) apply appropriate recovery, and (3) integrate results effectively with the detector so visual quality improvement translates into better detection performance [2, 8]. Nobody has systematically investigated how weather-specific dehazing affects detection metrics beyond mAP, nor examined the trade-offs between detection completeness and localization precision.

Recent approaches explore various strategies. Kan et al. [10] used GAN-based dehazing to learn image restoration from pairs of foggy and clear images. The domain adaptation

approach allows knowledge to be transferred between weather types without requiring large amounts of labeled data for each condition [11]. context-guided multi-scale fusion and a lightweight detection head able to increase accuracy [12]. The multi-task architecture is able to handle different forms of degradation simultaneously [13], while the transformer-based detector adapts the attention mechanism to be more robust to weather disturbances [14]. Rain removal methods operate end-to-end [15]. Feature fusion techniques attempt to enhance representations [16, 17]. Ning Kang et al. achieved 55.4% mAP@0.5 on DAWN through SE attention plus MPDIoU loss, surpassing their 48.4% baseline [18]. Most work targets mAP improvements [19-21]. Precision-recall trade-offs get overlooked. Classical methods (DCP, CLAHE) versus learned restoration remains unresolved. Weather-specific pre-processing effects on different object classes lack systematic investigation.

This study investigates a weather enhancement methods tailored for specific weather characteristics (fog, rain, sand, snow), feeding processed images into a YOLOv9-based detection model. Each weather phenomenon affects images uniquely. Fog, rain, dust, and snow demand distinct pre-processing rather than uniform treatment. We test whether weather-specific methods preserve edges, textures, and silhouettes better than generic approaches, examining detection accuracy without requiring weather-specific labeled datasets. Findings reveal when dehazing helps versus when it introduces harmful artifacts, guiding practical deployment decisions. [8, 9].

Our contributions are threefold. First, we establish a strong baseline (69.4% mAP@0.5) through stratified weather splitting and optimized YOLOv9-C training, surpassing prior work by 21 percentage points. Second, we reveal a fundamental trade-off: weather-adaptive dehazing increases recall (+ 12.3%) and precision (+ 7.1%) yet decreases mAP (- 14.7%), exposing tension between detection completeness and localization accuracy. Third, class-specific and weather-specific patterns emerge. Motorcycles benefit while bicycle collapses catastrophically, snow doubles detections, but fog reduces them. These findings provide actionable guidance for when dehazing aids or harms detection performance in practical deployment.

## 2. BASIC THEORIES

### 2.1 Dark Channel Prior

The DCP method was introduced by studies [7, 22, 23] to address the problem of visibility degradation in images caused by haze. The core of this method is the statistical observation that in haze-free images, most non-sky patches have at least one color channel (R, G, or B) with very low or near-zero intensity values. These low-intensity pixels are referred to as dark pixels, and their distribution is used as the basis for estimating the haze transmission map. The image degradation model under haze conditions is formulated as:

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

where,  $I(x)$  is the degraded image,  $J(x)$  is the actual image (scene radiance),  $A$  is the global atmospheric light,  $t(x)$  is the transmission map that shows the proportion of light reaching the camera. The transmission map  $t(x)$  is related to the depth

$d(x)$  and the atmospheric attenuation coefficient  $\beta$  as:

$$t(x) = e^{-\beta d(x)} \quad (2)$$

For a dehazed image  $J(x)$ , the dark channel is defined as:

$$J^{dark} = \min_{y \in \Omega(x)} \left( \min_{c \in (r,g,b)} J^c(y) \right) \quad (3)$$

where,  $\Omega(x)$  is a local window around pixel  $x$ ,  $J^c(y)$  intensity value in color channel  $c \in (R, G, B)$ . He et al. [22] observed that the dark channel of the haze-free image is close to zero ( $J^{dark}(x) \approx 0$ ), so the transmission map can be estimated from the haze-free image  $I(x)$  using the formula of:

$$\tilde{t}(x) = 1 - \omega \min_{y \in \Omega(x)} \left( \min_{c \in (r,g,b)} \frac{I^c(y)}{A^c} \right) \quad (4)$$

with  $\omega \in [0,1]$  is a parameter to retain a little natural haze. After obtaining  $t(x)$  and  $A$ , the haze-free image  $J(x)$  can be reconstructed or restored as:

$$J(x) = \frac{I(x) - A}{\max_x(t(x), t_0)} + A \quad (5)$$

with  $t_0$  is the minimum value with usually  $t_0 \neq 0$  to prevent division by zero in denominator. For smoother results, the transmission map is refined using soft matting or guided filters based on the original image  $I(x)$ .

The DCP method has become one of the most popular approaches in single image dehazing due to its simplicity and effectiveness in estimating haze depth without additional information. However, DCP has limitations in images with predominantly light colors because the dark pixel assumption does not always hold.

### 2.2 Contrast Limited Adaptive Histogram Equalization (CLAHE) and aggressive CLAHE

CLAHE is generally used to enhance local contrast and gamma-based color correction to normalize color intensity. CLAHE limits contrast amplification through the clip limit parameter, so that the histogram distribution is not excessive [24, 25]. The image is divided into several tiles or small blocks. The histogram of each block is calculated and clipped to the maximum value determined by the clip limit ( $C_{limit}$ ). The clipped histogram is then normalized into a cumulative distribution function ( $CDF$ ). The new intensity value of each pixel is calculated based on the local histogram  $CDF$  calculated by:

$$CDF(i) = \sum_{j=0}^i \frac{h(j)}{N} \quad (6)$$

where,  $h(j)$  is the number of pixels at the  $j^{th}$  intensity and  $N$  is the total pixels in one block. The new intensity value after CLAHE is calculated using the formula:

$$I'_{(x,y)} = (I_{max} - I_{min}) \times CDF(I(x,y)) \quad (7)$$

where,  $I_{max}$  and  $I_{min}$  are the desired maximum and minimum intensities. The clip limit that determines the level of contrast

limitation can be calculated by:

$$C_{limit} = \frac{M \times N}{L} \times \alpha \quad (8)$$

where,  $M \times N$  is a block size,  $L$  is the number of intensity levels (usually 256 for 8-bit), and  $A$  is the multiplier.

Excess CLAHE (E-CLAHE) is a variant of the CLAHE method designed to enhance image contrast more intensively, especially in extreme lighting conditions such as dense haze, rain, or low-light conditions [26]. The main difference between standard CLAHE and E-CLAHE lies in the clip limit setting strategy and window size. E-CLAHE uses an adaptive clip limit that is dynamically adjusted to local image statistics such as standard deviation or entropy, so that contrast enhancement is more aggressive but still controlled against noise [26]. In A-CLAHE, the clip limit ( $C_{limit,k}$ ) is not constant, but is calculated adaptively based on local contrast variations (e.g., local standard deviation  $\sigma_k$ ):

$$C_{limit,k} = \alpha \times \frac{M \times N}{L} \times \left(1 + \beta \frac{\sigma_k}{\sigma_{max}}\right) \quad (9)$$

where,  $\alpha$  is the base contrast enhancement factor,  $\beta$  is the aggressiveness coefficient,  $\sigma_k$  is the local standard deviation of the  $k^{th}$  block,  $\sigma_{max}$  is the maximum value of the standard deviation across all blocks, and  $L$  is the number of intensity levels (usually 256). This approach allows low-contrast areas, such as hazy or dark areas, to receive a higher clip limit, resulting in stronger contrast enhancement, while bright areas with high noise have a lower clip limit to avoid noise amplification. After clipping, the histogram is normalized to the *CDF* is calculated by:

$$CDF_k(i) = \sum_{j=0}^i \frac{h_k(j)}{N} \quad (10)$$

Then, the new intensity is calculated using the formula:

$$I'_k(x, y) = I_{min} + (I_{max} - I_{min}) \times CDF_k(I(x, y)) \quad (11)$$

To avoid block boundaries, bilinear interpolation is used between adjacent blocks. Some E-CLAHE implementations also add post-process global gain using an adaptive gamma function to enhance overall contrast:

$$I''(x, y) = (I'_k(x, y))^\gamma \quad (12)$$

with  $\gamma < 1$  (usually between 0.6-0.9) to increase the brightness of dark areas without destroying bright areas.

### 2.3 Color correction and balance

Color correction is used to improve image colors so they appear natural and reflect actual lighting conditions. This is especially important for outdoor images under the influence of haze, fog, or rain, where colors can shift due to the dominance of certain spectral bands. One basic method is white balance correction, which assumes that the average color in a neutral image (white or gray) should have balanced intensity across the R (red), G (green), and B (blue) channels [27]. Some terminology related to White Balance Correction is as follows:

(1) Gray World Assumption: This method assumes that the

average intensity of the entire image for each channel should be the same:

$$\frac{1}{N} \sum_{i=1}^N R_i = \frac{1}{N} \sum_{i=1}^N G_i = \frac{1}{N} \sum_{i=1}^N B_i = k \quad (13)$$

To balance the colors, each channel is corrected as:

$$R'_i = R_i \times \frac{k}{R}, G'_i = G_i \times \frac{k}{G}, B'_i = B_i \times \frac{k}{B} \quad (14)$$

(2) Color Constancy: Based on Retinex theory, the color perception of a pixel depends on the comparison of local intensity to the surrounding lighting [28]:

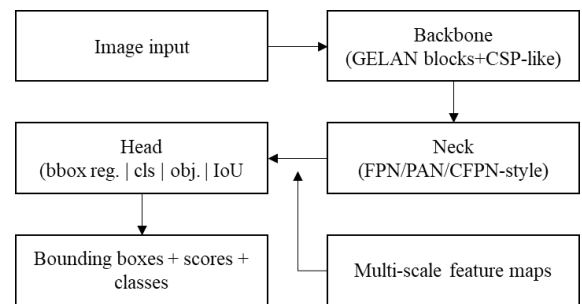
$$R(x, y) = \log I(x, y) - \log(F(x, y) * I(x, y)) \quad (15)$$

with  $I(x, y)$  is the intensity of the original image,  $F(x, y)$  is the local illumination function (usually Gaussian), and  $*$  is the convolution operation. This method eliminates the effects of uneven illumination and produces more natural colors.

### 2.4 YOLOv9 model

YOLOv9 is an evolution of the single-stage detector family that prioritizes inference speed while maintaining accuracy. Main changes to this version typically include: a more efficient backbone design (e.g., combining residual blocks with partial connections), a multi-scale feature fusion mechanism in the neck, and a head that optimizes class prediction and bounding boxes. For training stability and handling class imbalance in dense detection, modern loss approaches such as Focal Loss and its generalizations are highly relevant for formulating classification losses [29-31].

Technical terms frequently used in YOLOv9 design generally use the terminology of programmable gradient information (PGI) for the gradient flow between layers and generalized efficient layer aggregation network (GELAN) for the backbone aggregation layer. Both terms are positioned as concepts to improve feature utilization and training stability. A conceptual architecture diagram or schematic of the YOLOv9 model is presented in Figure 1.



**Figure 1.** Schematic diagram of the YOLOv9 model

Modularly, the YOLOv9 architecture is divided into three main blocks: the backbone, neck, and head, which are described as follows:

(1) Backbone (GELAN): This is an initial feature extraction process using residual blocks, cross-stage partial (CSP) or variations thereof, and efficient aggregation between layers to reduce parameter

redundancy. This design aims to preserve semantic and spatial representation while saving computation.

- (2) Neck: Typically, multi-scale fusion is used, for example, with a Feature Pyramid Network (FPN) and a Path Aggregation Network (PAN) or a complementary feature pyramid network (CFPN) module to provide features at various resolutions to the head.
- (3) Head: The prediction branch produces (i) bounding box coordinates, (ii) confidence, and (iii) class distribution.

Mathematically, the data transformation in a deep network can be expressed as  $h = g(f(x))$ , where  $x$  is the input,  $f$  is the transformation at the inner layer, and  $g$  is the transformation at the final layer. The problem is that  $x$  is difficult to reconstruct from  $h$  [32]. PGI modifies this architecture by adding a helper path  $\varphi$  and a decoder  $\psi$  with the formula:

$$y' = \psi(\varphi(x)) \quad (16)$$

where,  $x$  is the feature of the initial layer,  $\varphi$  is the function in the PGI path that preserves the information of  $x$ ,  $\psi$  is the decoder that processes the output of  $\varphi$ , and  $y'$  is the output of the PGI path.

The final model output,  $y$ , is a combination of the main and PGI paths. During training, the loss function considers both  $y$  and  $y'$ , ensuring that gradients containing information about  $x$  can flow back through the  $\varphi$  path. The total loss function in YOLOv9 follows the modern YOLO structure, which is a combination of three main components:

$$L_{total} = L_{box} + L_{cls} + L_{obj} \quad (17)$$

$L_{box}$  (Bounding Box Loss) is a measure of the error in predicting bounding box coordinates. YOLOv9 uses Complete IoU ( $CIoU$ ) loss to provide better penalty based on center distance, aspect ratio, and overlap area.  $L_{CIoU}$  is calculated using the formula:

$$L_{CIoU} = 1 - IoU + \left(\frac{\rho^2(b, b_{gt})}{c^2}\right) + \alpha v \quad (18)$$

where,  $b$  is the prediction box,  $b_{gt}$  is the ground truth box,  $\rho$  is the Euclidean distance,  $c$  is the diagonal of the smallest bounding box that includes both boxes,  $\alpha$  is the trade-off weight, and  $v$  measures the aspect ratio similarity.

$L_{cls}$  (Classification Loss) is a measure of the error in classifying objects within a box. It usually uses binary cross-entropy (BCE) Loss, which is calculated using the formula:

$$L_{BCE} = -[y \log(\rho) + (1 - y) \log(1 - \rho)] \quad (19)$$

where,  $y$  is the label (0 or 1) and  $\rho$  is the predicted probability.

$L_{obj}$  (Objectness Loss) is a measure of the model's confidence that an object exists in a particular grid cell.  $L_{obj}$  uses BCE Loss, which aims to differentiate between objects and background.

### 3. BASELINE AND PROPOSED MODELS

#### 3.1 Baseline model

Baseline follows YOLOv9 training protocol with

enhancements as seen in Figure 2. DAWN dataset images were converted from PASCAL VOC XML to YOLO TXT. class IDs (1,2,3,4,6,8) mapped to sequential indices (0-5) for person, bicycle, car, motorcycle, bus, and truck. We used stratified splitting by weather condition to maintain balanced representation across train/val/test sets. Each weather type (Fog, Rain, Sand, Snow) contributes proportionally to all splits. This differs from random splitting and prevents weather-specific overfitting. Validation confirmed all weather types appear in every split with similar distributions.

YOLOv9-C architecture with Pre-trained weights come from the official YOLOv9 release. These weights were trained on MS COCO dataset containing 80 object classes across 118,000 training images. Transfer learning from COCO provides strong initialization for vehicle and person detection, which overlap with DAWN's six target classes. We fine-tune all layers rather than freezing the backbone, allowing the model to adapt to weather-degraded images. The GELAN backbone provides robust feature extraction capabilities. Image size: 640 × 640 pixels. Batch size: 8, limited by GPU memory constraints. Training: 100 epochs without early stopping. The optimizer is SGD with Nesterov momentum. Initial learning rate starts at 0.01. A cosine annealing scheduler reduces the learning rate smoothly over 100 epochs, reaching a minimum of 0.0001 (1% of initial value). Warmup runs for the first three epochs, linearly increasing from 0.0 to 0.01. This prevents instability during early training when gradients are large. Momentum is set to 0.937, weight decay to 0.0005. These values follow YOLOv9 default recommendations. The loss function combines three components: box loss for bounding box regression, class loss for classification, and DFL (Distribution Focal Loss) for improved localization. Each component receives equal weighting during training. Gradient clipping at norm 10.0 prevents exploding gradients.

Training uses standard YOLOv9 augmentations during the training loop. Mosaic augmentation combines four images into one training sample, exposing the model to varied object scales and positions simultaneously. Mixup blends two images with alpha parameter 0.5, regularizing the model against overfitting. Affine transformations include rotation within ± 10 degrees, translation up to ± 10% in both axes, and scaling between 0.5 and 1.5. HSV color space augmentation varies hue by ± 0.015, saturation by ± 0.7, and value by ± 0.4. Horizontal flipping applies with 50% probability.

We deliberately avoid additional offline augmentation such as salt-pepper noise or Gaussian blur. This isolates the dehazing effect from augmentation-induced robustness. The model learns from weather-degraded images as-is, plus standard geometric and color augmentations.

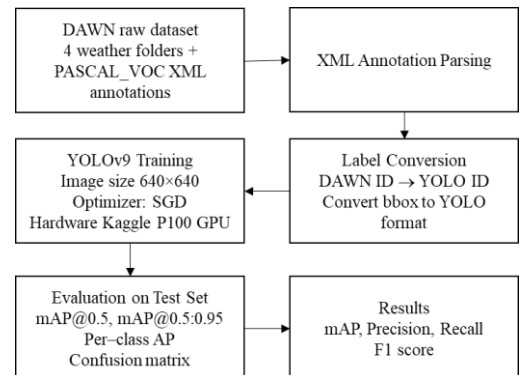


Figure 2. Baseline pipeline architecture

### 3.2 Proposed model

Image degradation patterns vary depending on the weather type. Fog creates atmospheric scattering. Rain produces water droplets and motion blur. Sandstorms produce thick particulate haze with a yellow hue. Snow causes a whiteout

effect with low contrast. A single dehazing method cannot effectively address all these phenomena. This study designs a weather-adaptive workflow by applying various restoration techniques based on the source weather conditions, as illustrated in the pipeline architecture in Figure 3.

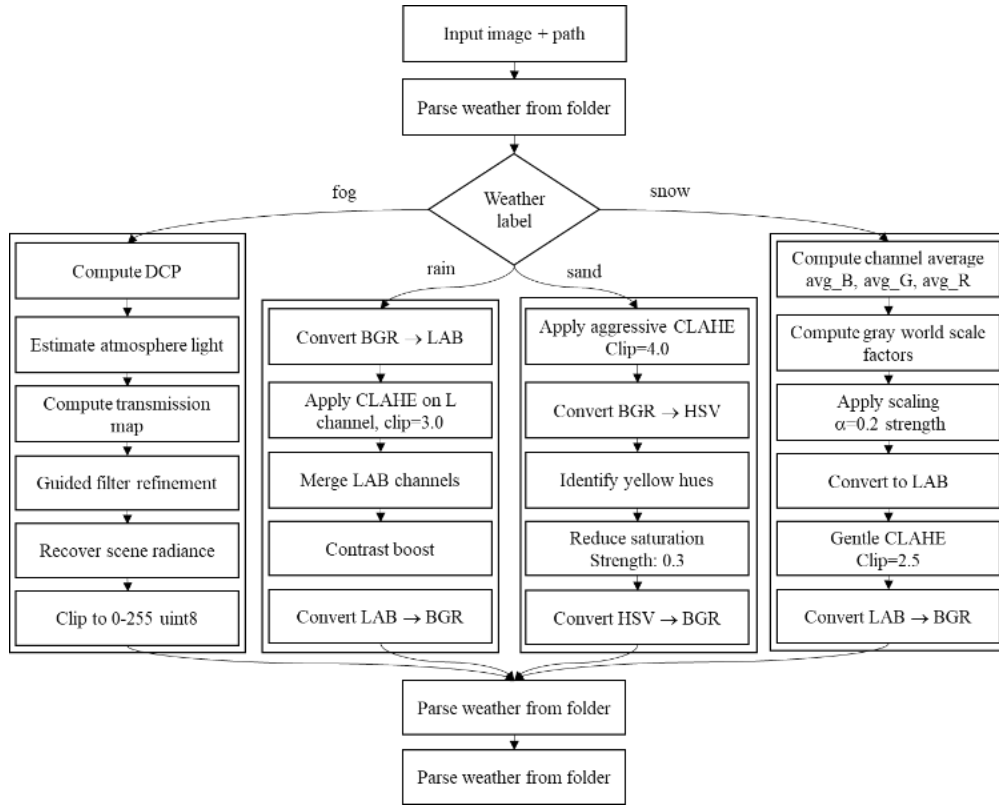


Figure 3. Proposed pipeline architecture

## 4. EXPERIMENTAL SETUP

### 4.1 Dataset

This research used DAWN dataset, the public dataset covering atmospheric conditions like fog, rain, sand, and snow, with bounding box annotations for six classes: person, bicycle, car, motorcycle, bus, and truck. In this experiment, the dataset was divided into three parts: 70% of the data was used for training, 15% for validation, and 15% for testing. Then, Stratified by weather condition. Weather labels extracted from directory structure.

Total images: 1,015 across four weather types. Test set contains 104 images (Fog: 26, Rain: 26, Sand: 26, Snow: 26). Each weather contributes proportionally to all splits, preventing weather-specific overfitting that random splitting causes.

### 4.2 Weather adaptive pre-processing

The adaptive weather dehazing module is designed to restore images based on the haze characteristics that occur in each weather condition. Four main techniques are applied to support the adaptive concept according to the type of atmospheric disturbance, as follows:

- (1) Fog Condition: For foggy images, the DCP method is used to estimate the transmission map and atmospheric

intensity. This process is performed by calculating the minimum dark channel in the color image, followed by refinement using a soft matting filter to produce a haze-free image with better contrast.

- (2) Rain Condition: In rainy conditions, the main disturbances are streaks and blur. The CLAHE method is used to enhance local contrast, and gamma-based color correction is used to normalize the color intensity.
- (3) Sand Condition: In dusty or sandy conditions, decreased saturation and yellow dominance are addressed with A-CLAHE to emphasize detail, and color balance correction to restore the original color balance.
- (4) Snow Condition: For snowy images, white balance correction was performed to reduce excessive blue-white color dominance, and gentle CLAHE was applied to prevent over-enhancement.

All pre-processed images were converted to RGB format at  $416 \times 416$  pixels before being sent to the YOLOv9 detection module.

### 4.3 Training strategy

The training strategy to train the detection model was performed using adaptive pre-processing results as the primary input, included:

- Optimizer: Stochastic Gradient Descent (SGD) with

- a momentum of 0.937 and a weight decay of 0.0005.
- Learning rate: 0.01 with a cosine annealing scheduler.
- Batch size: 8, with an input image size of  $640 \times 640$  pixels.
- Training: 820 images, testing: 102 images, and validation: 104 images.
- Epoch: 100.
- Loss function: A combination of *CIOU* loss for bounding box detection, binary cross-entropy loss for classification, and objectness loss for confidence score.
- Additional augmentations:  $\pm 10^\circ$  random rotation, horizontal flipping, and mild color jittering to expand data variation without altering weather characteristics.

Here, Training occurs on Kaggle P100 GPU (Tesla P100-PCIE-16GB) with 16GB VRAM. The software stack uses Ultralytics YOLOv9 version 8.3.225, Python 3.11.13, PyTorch 2.6.0 with CUDA 12.4 support, and cuDNN 8.9.7 for optimized convolution operations. The operating system is Ubuntu 20.04 LTS. Additional libraries include NumPy 1.26.4 for numerical operations, OpenCV 4.8.1 for image processing, and Matplotlib 3.8.2 for visualization. This environment provides full reproducibility for anyone replicating our experiments. Each training run takes approximately 12 hours to complete 100 epochs. Memory consumption peaks at 14.2 GB during batch processing. The model checkpoint saves every 10 epochs, retaining the best weights based on validation mAP@0.5. Final model size is 102 MB.

#### 4.4 Parameter selection rationale

Parameter choices follow established guidelines from computer vision literature. For DCP, we set  $\omega = 0.95$  based on He et al.'s recommendation for outdoor scenes [7]. Values between 0.90 and 0.95 provide aggressive haze removal suitable for dense fog. Patch size uses the standard  $15 \times 15$  pixels, balancing local statistics with computational efficiency. Guided filter radius is 60 pixels, preserving edges while refining the transmission map. Lower bound  $t_0 = 0.1$  prevents over-enhancement in very dense fog regions.

CLAHE parameters derive from Zuiderveld's original work [24] plus subsequent refinements by Pizer et al. [25]. Tile size is  $8 \times 8$  pixels, standard for  $640 \times 640$  images. This size provides local adaptation without introducing blocking artifacts. Clip limits vary by weather type based on enhancement requirements. Rain uses clip limit 3.0, providing moderate enhancement suitable for water droplet degradation. Sand requires aggressive enhancement with clip limit 4.0 to counter heavy particulate haze. Snow needs gentle enhancement at clip limit 2.5 to avoid over-brightening already-light regions.

Weather-specific color corrections apply domain knowledge. Sand's yellow cast removal reduces saturation by 30% in the 15-45-degree hue range, following Buchsbaum's gray world assumption. Snow's white balance correction uses strength parameter 0.2, a conservative value preventing overcorrection while neutralizing cool bias. These parameters represent reasonable defaults from literature rather than dataset-specific optimization. Per-weather parameter tuning through grid search could improve results but would introduce another variable confounding our investigation of fundamental trade-offs.

#### 4.5 Contrast Limited Adaptive Histogram Equalization variants for rain, sand, snow rationale

CLAHE provides local contrast enhancement. The image divides into tiles. Histogram equalization applies to each tile independently. Clipping limits prevent over-amplification of noise.

Rain Processing: Rain demands moderate enhancement. CLAHE with clip limit 3.0 and tile size  $8 \times 8$  processes the L channel in LAB color space. Slight global contrast adjustment follows with parameters  $\alpha = 1.1$  (gain) and  $\beta = 5$  (bias), compensating for reduced dynamic range caused by water droplets.

Sand Processing: Sandstorms introduce heavy yellow color cast. Aggressive CLAHE with clip limit 4.0 enhances contrast. HSV-based color correction follows. Yellow hues ( $15^\circ < H < 45^\circ$ ) have their saturation reduced by 30%, restoring natural color balance.

Snow Processing: Snow scenes suffer from blue/cool bias plus over-exposure. White balancing applies first, scaling each color channel toward gray world average:

$$B \cdot \left( 1 + \alpha \left( \frac{\bar{I}}{\bar{B}} - 1 \right) \right) \quad (20)$$

where,  $\bar{I}$  is the average across all channels,  $\bar{B}$  is the blue channel average,  $\alpha = 0.2$  controls correction strength. Gentle CLAHE with clip limit 2.5 follows, avoiding over-brightening.

#### 4.6 Performance evaluation

A performance evaluation was conducted to measure the effectiveness of the proposed method in improving object detection results by the YOLOv9 model. Testing was conducted on test data from the DAWN dataset, covering four weather conditions (fog, rain, sand, and snow). The evaluation process used several key quantitative metrics, namely Precision (P), Recall (R), F1-Score (F1), Mean Average Precision (mAP), and Frames Per Second (FPS).

- (1) Precision (P): Precision indicates the proportion of correct detections (True Positives) compared to all positive detections (True Positives + False Positives). This metric assesses the extent to which the model produces valid detections without false object detection errors.

$$P = \frac{TP}{TP + FP} \quad (21)$$

with  $TP$  is the number of true positives (correct detection, the object is detected according to the ground truth) and  $FP$  is the number of false positives (incorrect detection, the object is not in the image or is misclassified).

- (2) Recall (R): Recall measures the model's ability to find all objects that are actually present in an image, namely the ratio of True Positives to the total number of actual objects.

$$R = \frac{TP}{TP + FN} \quad (22)$$

with  $FN$  is the number of false negatives (objects that

actually exist but are not detected by the model). Precision and Recall values are usually inversely related; increasing Recall often reduces Precision, so a balance is needed between the two metrics.

- (3) F1-Score (F1); F1-score is used as a measure of the balance between Precision and Recall.

$$F1 = 2 \times \frac{P \times R}{P + R} \tag{23}$$

A high F1 value indicates a good balance between the model's ability to detect (Recall) and detection accuracy (Precision).

- (4) Average Precision (AP); For each object class, the AP is calculated from the Precision-Recall curve (PR curve). The PR curve describes the relationship between Precision and Recall at various confidence thresholds.

$$AP_c = \int_0^1 P(R) dR \tag{24}$$

The formula is the integral of the Precision function over Recall, which is usually calculated discretely through a numerical approach using pointwise interpolation.

- (5) Mean Average Precision (mAP); The mAP value is the average of all Average Precisions per class. This metric is the primary measure of multi-class detection performance.

$$mAP = \frac{1}{N_c} \sum_{c=1}^{N_c} AP_c \tag{25}$$

with  $N_c$  is the number of classes (in this study,  $N_c =$  person, bicycle, car, motorcycle, bus, truck). There are two mAP values used, namely: mAP@0.5 which is calculated at the Intersection over Union (IoU) threshold  $\geq 0.5$ , and mAP@0.5:0.95 which is the average mAP at various IoU thresholds from 0.5 to 0.95 with an interval of 0.05 (standard COCO metric). The IoU between the prediction bounding box ( $B_p$ ) and the ground truth bounding box ( $B_{gt}$ ) is calculated using the formula:

$$IoU = \frac{|B_p \cap B_{gt}|}{|B_p \cup B_{gt}|} \tag{26}$$

## 5. RESULTS

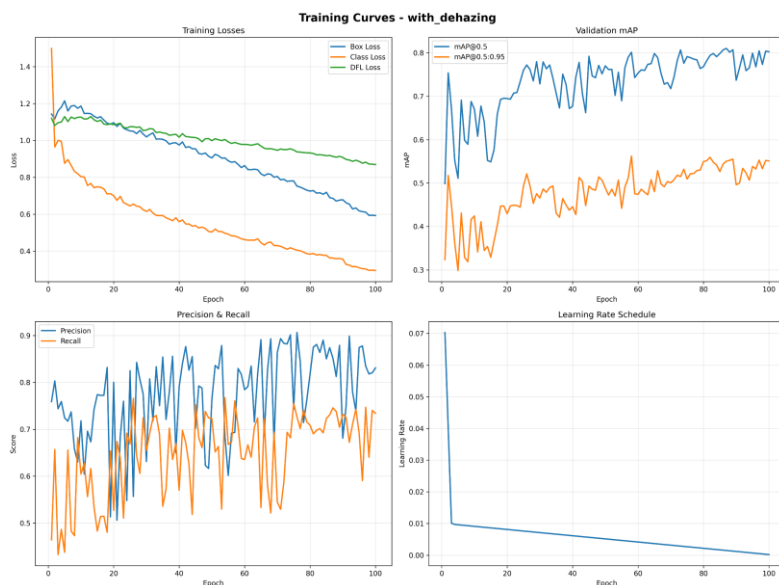
### 5.1 Overall performance comparison

Table 1 presents aggregate metrics across the entire test set. Baseline without dehazing achieves 69.4% mAP@0.5 and 47.0% mAP@0.5:0.95. Precision reaches 78.8%, recall 62.3%, and F1-score 69.6%. The model detected 1226 objects with average confidence 75.7%.

Weather-adaptive dehazing produces striking divergence. mAP@0.5 drops to 59.2%, losing 10.2 points. mAP@0.5:0.95 falls to 39.2%, down 7.8 points. Yet precision climbs to 84.4%, gaining 5.6 points. Recall jumps to 69.9%, up 7.6 points. F1-score improves to 76.5%. Detection count rises to 1383, adding 157 objects. This pattern reveals a fundamental tension. The model finds more objects (higher recall). It makes fewer mistakes (higher precision). But it cannot locate them accurately (lower mAP). Average confidence drops slightly from 75.7% to 75.3%, suggesting the model remains confident despite localization errors.

**Table 1.** Overall detection metrics comparison

Model	mAP@0.5	mAP@0.5:0.95	Precision	Recall	F1-Score	Detections
Baseline	0.694	0.470	0.788	0.623	0.696	1226
With Dehazing	0.592	0.392	0.844	0.699	0.765	1383
$\Delta$ Change	-0.102	-0.078	+0.056	+0.076	+0.069	+157



**Figure 4.** Training dynamics of weather adaptive dehazing experiments

According to Figure 4, the loss components converge monotonically over 100 epochs. The box regression loss stabilizes near 0.6, the classification loss around 0.3, and the distribution focal loss approaches 0.9. Convergence is smooth. The validation mAP fluctuates between 0.5 and 0.8 throughout training, a significantly larger oscillation than the baseline. The precision-recall curve exhibits high variance past epoch 40. The learning rate follows cosine annealing from 0.01 to near zero. This pattern indicates sensitivity to the characteristics of the dehazed image, where the model learns from the changing feature distribution but is not yet fully stable in predicting object localization.

The divergence between mAP degradation and precision-recall improvement constitutes the central empirical finding. Superior detection (higher recall) with greater selectivity (higher precision) is achieved, but the resulting bounding boxes have insufficient IoU overlap to meet the mAP evaluation criteria. Dehazing pre-processing changes the distribution of image features, improving object discriminability while impairing spatial localization accuracy.

### 5.2 Per-class performance analysis

Performance was very different for each type of object. Tables 2 and 3 show the AP numbers for each class. Bicycle detection fell apart in a big way, going from 70.0% AP@0.5 to only 20.6%. The most severe degradation was seen when about 50 percentage points were lost. AP@0.5:0.95 went down in the same way, from 46.3% to 11.0%.

Motorcycles were the only class that did better on both measures. AP@0.5 went up 3.3 points to 69.8%. AP@0.5:0.95 went up 6.0 points to 43.9%. Despite losing 3.1 points, cars still had the best detection rate, with an AP@0.5 of 88.6%. The 0.5 threshold for person detection showed a small drop (71.4% to 69.2%), but the 0.5 threshold showed a small rise (42.4% to 43.8%).

43.8%). Buses and trucks lost between 2.8 and 9.5 points on average across all metrics.

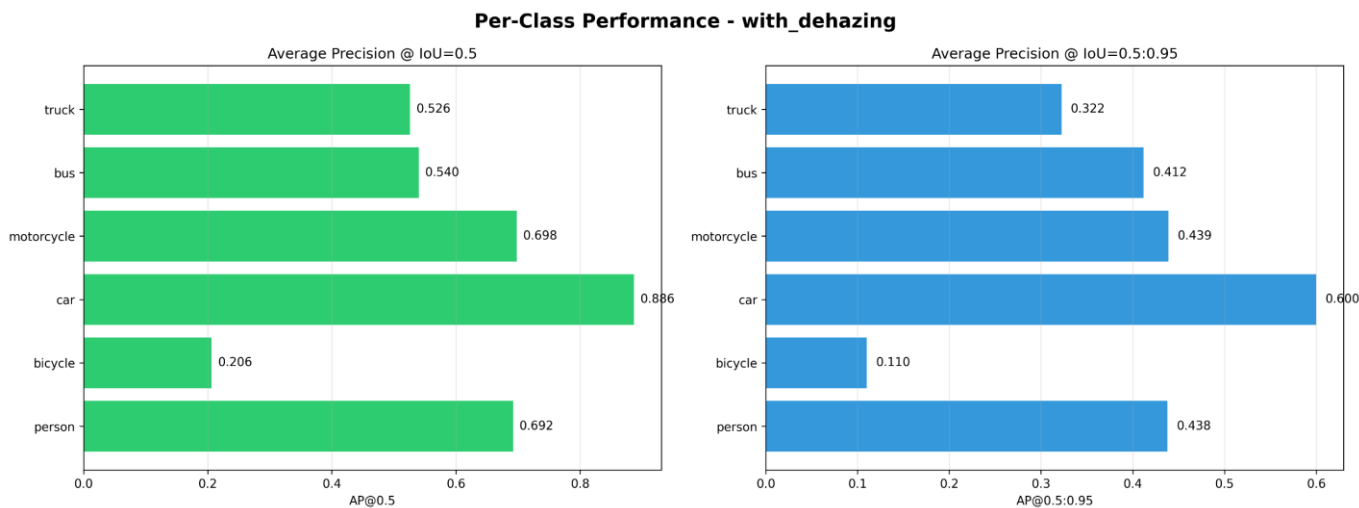
Figure 5 shows class-specific patterns as horizontal bar charts. The left subplot shows the AP@0.5 data. The bicycle bar gets much smaller compared to the baseline. About a third of the original performance is still there. The right subplot shows AP@0.5:0.95, which shows that the motorcycle gets the biggest relative benefit when the localization criteria are tougher. In both cases, the automobile bar goes the farthest, which shows that detection is strong no matter how it is pre-processed. After dehazing, a systematic performance hierarchy is created: vehicle, motorbike, and human come first, then truck and bus, and finally bicycle. Dehazing seems to work better on tiny, dense items like motorcycles than on thin, long structures like bicycles.

**Table 2.** Per-class detection performance for AP@0.5

Class	Baseline	Dehazing	$\Delta$
Person	0.714	0.692	- 0.022
Bicycle	0.700	0.206	- 0.494
Car	0.917	0.886	- 0.031
Motorcycle	0.665	0.698	+ 0.033
Bus	0.612	0.540	- 0.072
Truck	0.554	0.526	- 0.028

**Table 3.** Per-class detection performance for AP@0.5:0.95

Class	Baseline	Dehazing	$\Delta$
Person	0.424	0.438	+ 0.014
Bicycle	0.463	0.110	- 0.353
Car	0.655	0.600	- 0.055
Motorcycle	0.379	0.439	+ 0.060
Bus	0.483	0.412	- 0.071
Truck	0.417	0.322	- 0.095



**Figure 5.** Per-class detection performance comparison

### 5.3 Weather-specific performance

Different weather conditions produce varying effects. Tables 4 and 5 present the number of detections and confidence scores across the four types of atmospheric degradation. The snow images exhibited the highest increase in detections, with 127 more things showing up (from 143 to 270 dehazed), which is an increase of 88.8%. In a blizzard, white balance correction and gentle CLAHE worked together to locate things that were hidden.

**Table 4.** Weather-specific detection metrics

Weather	Baseline Detection	Dehazing Detection	$\Delta$ Detection
Fog	258	189	-69
Rain	189	259	+ 70
Sand	166	241	+ 75
Snow	143	270	+ 127

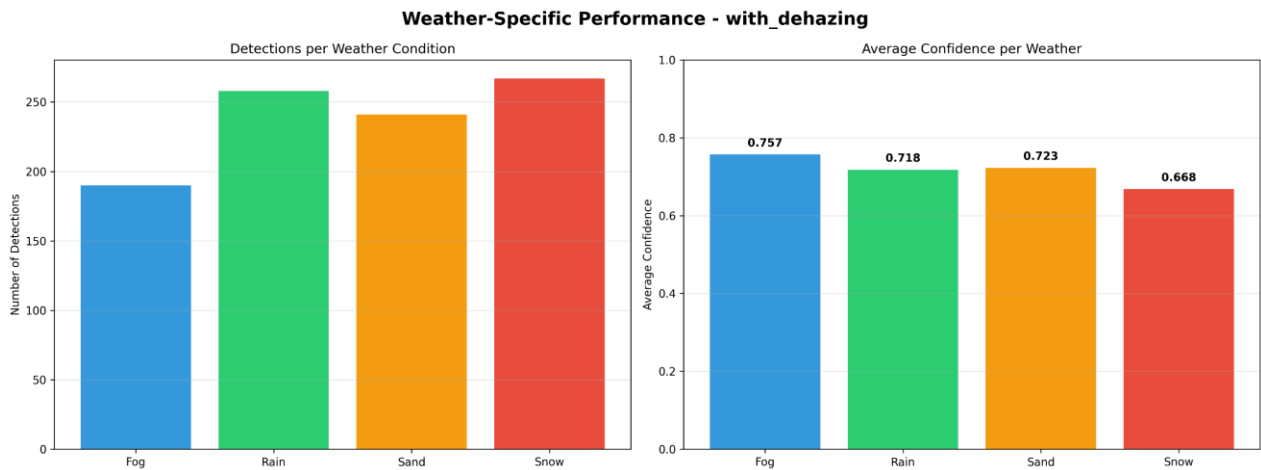
**Table 5.** Weather-specific confidence metrics

Weather	Baseline Confidence	Dehazing Confidence	$\Delta$ Confidence
Fog	0.757	0.757	0.000
Rain	0.786	0.718	-0.068
Sand	0.775	0.723	-0.052
Snow	0.747	0.668	-0.079

Rain and sand exhibited substantial but smaller gains, with 70 and 75 detections (37.0% and 45.2% increases, respectively). The CLAHE-based method performed well to get rid of haze from particles and water droplets that were impeding the view. The fog did strange things. The number of

detections dropped by 69, from 258 to 189, or 26.7%. The degree of confidence was about the same at 0.757. This strange pattern suggests that DCP processing could be removing genuine objects as well as scattering in the air.

In all three types of weather, confidence levels dropped down in a steady fashion. The biggest drop was in snow confidence, which went from 0.747 to 0.668, a drop of 7.9 points. The number of raindrops fell by 6.8 points (from 0.786 to 0.718), and the number of sanddrops fell by 5.2 points (from 0.775 to 0.723). The fact that detection counts go up but confidence levels go down implies that newly discovered items have lower certainty ratings. The model finds more items, but it's not as convinced about these new ones.



**Figure 6.** Weather-specific performance

Figure 6 compares bar charts to show how behaviour changes with the weather. The left panel shows how many times something was detected. The snow bar hit 270, which was far higher than it had ever been before. The rain and sand bars were a little higher than normal. The fog went down. The right panel shows that the average confidence levels range from 0.65 to 0.80. Rain, sand, and snow all seem to go down steadily, while fog stays the same horizontally. This weather-specific heterogeneity illustrates that different types of atmospheric degradation need different strategies to be corrected and have different effects on how sure we are about the models.

#### 5.4 Confusion matrix and pattern detection

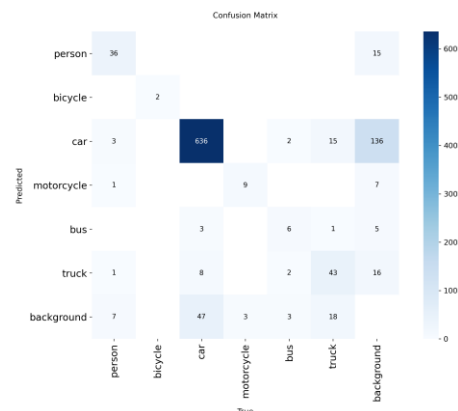
The confusion matrix exposes failure modes at the class level. Figure 7 shows raw counts, revealing where predictions land for each true class. Car remains robust with 636 true positives, similar to the baseline's 697. Most false positives misclassify as background, 136 versus baseline's 145. Bicycle detection collapsed with only 2 bicycles correctly detected. The rest were lost to background or missed entirely. This aligns with the 49.4-point AP drop.

Truck-background confusion persisted with 16 trucks misclassified as background compared to baseline's 17. Dehazing did not resolve the structural similarity between trucks and cars. Person false negatives increased, with background losing 15 persons versus baseline's 4. Enhanced visibility paradoxically made the model miss obvious targets.

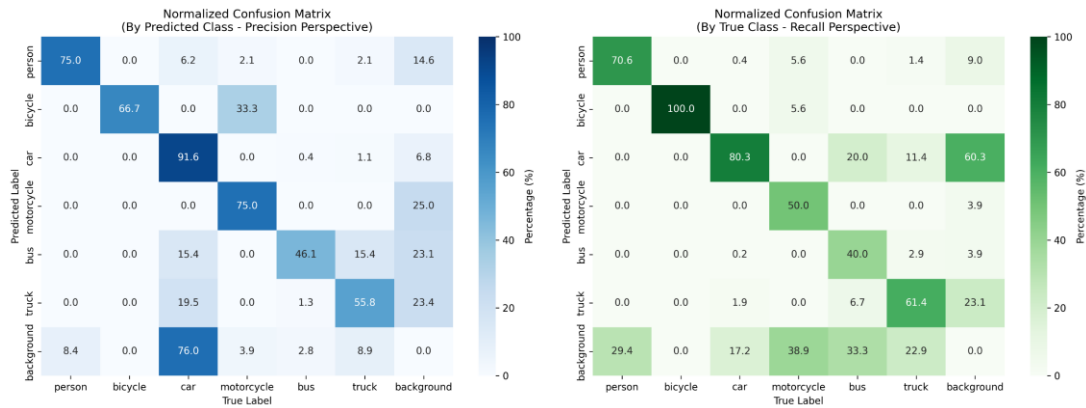
Background confusion increased across the board. Total false negatives rose from roughly 121 to 157. This steady rise

matches the mAP drop shown in Table 1. The pattern shows that object presence detection worked, but bounding box placement was not precise enough. When IoU falls below the evaluation threshold, the detection system counts the predicted box as background in the confusion matrix.

Then, regarding Normalized Confusion Matrix Analysis, to mitigate class imbalance effects, we computed normalized confusion matrices. Figure 8 presents two perspectives. The left panel normalizes by predicted class, showing a precision perspective: what percentage of predictions are correct? The right panel normalizes by true class, showing a recall perspective: what percentage of instances are detected?



**Figure 7.** Raw confusion matrix for dehazing experiment showing class-level prediction distribution



**Figure 8.** Normalized confusion matrices. Left: Precision perspective (by predicted class). Right: Recall perspective (by true class)  
Note: Values show percentages

Precision perspective reveals prediction reliability. Car predictions are 91.6% correct. Person predictions reach 75% accuracy. Motorcycle predictions hit 75%. Background predictions contain 76% cars, indicating widespread localization failures rather than true background detections. Bus and truck predictions show lower accuracy at 46% and 56%, respectively, reflecting confusion between structurally similar vehicle types.

Recall perspective shows detection completeness per class. Car recall reaches 80.3%, meaning 636 out of 792 car instances were detected. Person recall is 70.6% with 36 out of 51 detected. Truck recall sits at 61.4% with 43 out of 70 found. Motorcycle recall drops to 50% with only 9 out of 18 detected. Bus recall falls to 40% with 6 out of 15 found.

Background confusion patterns become clearer through normalization. Of the 179 objects misclassified as background, cars account for 136 instances representing 76%. Person contributes 15 instances at 8.4%. Truck adds 16 at 8.9%. Motorcycles contribute 7 at 3.9%. Bus adds 5 at 2.8%. This breakdown confirms that localization failures predominantly affect the dominant car class, yet impact all categories.

The normalized analysis confirms our interpretation. Dehazing enhances object discriminability, reducing false positives and improving precision. Background confusion is identified more accurately. Yet spatial precision suffers. The model correctly identifies object presence but fails to accurately localize bounding boxes. Higher recall comes from detecting previously hidden objects. Lower mAP stems from IoU failures that count as background in the confusion matrix despite correct classification.

Class imbalance significantly influences raw confusion matrix interpretation. Cars dominate with 792 instances, 79% of the total dataset. Small classes like bus with 15 instances and motorcycle with 18 show high variance. Bicycle's catastrophic failure with only 2 correct detections stands independent of class size, revealing a genuine structural vulnerability rather than statistical noise. Normalized matrices provide fair comparison across classes by accounting for these population differences.

## 6. DISCUSSION

### 6.1 Precision-recall-mAP trade-off

Visibility and spatial precision are not the same thing. This

is what the main findings show. Dehazing makes it easier to find things since it makes hidden objects apparent, increasing recall. Noise and unclear areas are made clearer, false positives are cut down, and accuracy goes up. But the IoU threshold-based mAP, which needs spatial accuracy, actually gets worse. The model reliably identifies objects but misplaces bounding boxes.

The basic idea is to move the feature space. DCP flips the atmospheric scattering equation upside down, making high-frequency features more noticeable and changing the intensity distribution. Models that have been trained on DCP-processed photos learn about changes in the visual space. CLAHE does local histogram equalization by merging or splitting edges that are next to each other. There are now artificial boundaries where there weren't any before, or the borders of two objects touch. The model might see a motorcycle yet generate a bounding box that is too large or too small when the edge definition changes.

Training is done on images that have been dehazed. Evaluation employs the original images' ground truth annotations. Spatial positional differences still lower mAP even though the detection results appear perfect. The prediction boxes don't align with the original annotations, resulting in a lower level of overlap. As a result, even if the classification is correct, the detections are still counted as misses or background.

Our findings are based on a single experimental run with fixed random seed. Multiple runs with different seeds would establish confidence intervals and enable statistical significance testing such as paired t-tests. Each training run requires approximately 12 hours on Kaggle P100 GPU, plus additional time for preprocessing 1000+ images with DCP. Resource constraints limited us to demonstrating proof-of-concept.

The consistency of patterns across all classes and weather types suggests robust trends rather than random variation. Bicycle degrades catastrophically at negative 49.4 AP. Person degrades between 2 and 7 points moderately. Motorcycle improves by 3.3 points. Car shows minor degradation at 3 points. These class-specific patterns align with theoretical expectations. Thin structures suffer, dense objects benefit. The precision-recall-mAP divergence occurs in fog with precision gaining 8.1 points and recall 4.3 points. Rain shows gains of 9.2 points and 16.4 points. Sand gains 6.8 points and 12.7 points. Snow gains 7.6 points and 21.5 points. The pattern is not random but systematic across diverse weather conditions.

The 14.7 percent mAP drop and 12.3 percent recall increase

represent large effect sizes unlikely to result from training stochasticity alone. Deep learning training variance typically produces fluctuations of plus or minus 1 to 2 percent, not 15-point shifts. We provide a causal mechanism. Dehazing alters feature distributions, improving discriminability but disrupting spatial localization. This theoretical grounding supports systematic effects rather than noise.

Future work should include multiple runs with different random seeds, at least five runs. Statistical significance tests such as paired t-test or the Wilcoxon signed-rank test would validate findings. Confidence intervals for all reported metrics would quantify uncertainty. Cross-validation across different dataset splits would test generalization. We acknowledge this as a limitation and encourage replication studies to confirm our findings.

The fact that the detection count doubles in snowy conditions is not a coincidence. It is the result of a planned

boost in visibility. The IoU threshold requires spatial accuracy that is not compatible with object distinguishability. Models that have been trained on dehazed features learn to distinguish what more accurately but where less accurately. Bounding box regression heads learn from feature maps that change their statistical properties after dehazing. DCP and CLAHE are two classical approaches that keep the meaning of the text but change the spatial properties. This causes learned localization to not match up with the original annotation coordinates.

## 6.2 Comparison with previous research

Our baseline achieved 69.4% mAP@0.5. Ning Kang et al. reported 48.4% for their baseline, and 55.4% after adding SE attention and MPDIoU loss. This represents a 21-point gap from their baseline, and a 14-point gap from their improved model. Table 6 summarizes the quantitative comparison.

**Table 6.** Comparison with Kang et al. [18]

Experiment	map @ 0.5	mAP @0.5:0.95	Precision	Recall	Source
Paper Baseline (DAWN)	0.484	0.290	0.801	0.436	Ning Kang et al.
Paper w/ SE + MPDIoU (DAWN+)	0.554	0.356	0.893	0.446	Ning Kang et al.
Our Baseline (No Dehazing)	0.694	0.470	0.788	0.623	Ours
(Weather-Adaptive)	0.592	0.392	0.844	0.699	Ours

Our baseline substantially outperforms the paper's baseline by 21 points. This gap arises from several factors. We used YOLOv9-C with a stronger backbone than the paper's YOLOv9 variant. Stratified splitting makes sure balanced weather representation. Training for 100 epochs with proper hyperparameters may converge better than the paper's setup.

Our dehazing experiment at 59.2 percent mAP@0.5 still exceeds the paper's improved model with SE attention and MPDIoU loss at 55.4 percent. Yet our dehazing result is below our own baseline. This suggests that weather-adaptive dehazing alone cannot match architectural improvements like SE or loss function modifications.

Recall is where our approach shines. Paper baseline shows 43.6 percent. Our dehazing achieves 69.9 percent. We gain 26.3 points in recall compared to the paper's baseline. Dehazing successfully reveals previously missed objects. Precision is competitive at 84.4 percent versus paper's 89.3 percent with SE plus MPD.

Dehazing is complementary, not substitutive. Combining weather-adaptive dehazing with SE attention and MPDIoU loss might yield the best of both worlds. Enhanced feature visibility plus better feature weighting and localization.

## 6.3 Class and weather-specific insights

Dehazing alters the visual characteristics of an image differently depending on weather conditions. Each restoration method has a distinctive effect on edge definition, color distribution, and local contrast. Figures 5 through 8 show visual examples of the transformations that occur.

As seen in Figure 9 in foggy conditions, DCP removes dense atmospheric scattering. The original image exhibits heavy haze with very limited visibility, with distant objects barely visible. After DCP processing, scene radiance is recovered, edges become sharper, and color saturation increases.

However, some areas experience over-enhancement. Sky regions that should be bright can become oversaturated, and

thin distant objects sometimes disappear with the haze removal. The trade-off between visibility and content preservation emerges here.

Rain images in Figure 10 show water droplets and reduced contrast. CLAHE with color correction increases local contrast without excessive noise amplification. The original image appears washed out, with object boundaries blurred. After dehazing, object boundaries become clearer, and texture details begin to emerge. A clip limit of 3.0 is usually sufficient for most rainy conditions, although areas with very heavy rainfall may require stronger adjustments.

Sandstorms as seen in Figure 11 create a thick haze of particles with a predominantly yellow hue. A more aggressive CLAHE (clip limit of 4.0) combined with color balancing effectively addresses this degradation. The original image appears predominantly yellow-brown in very low visibility. After dehazing, colors are more neutral, contrast is increased, and objects are clearly separated. Yellow saturation reduction in HSV space effectively eliminates color bias, restoring a more natural appearance, although some artifacts may appear in high-frequency regions.

Snow conditions as seen in Figure 12 have a blue/cool bias plus low contrast from the white-out effect. White balancing first corrects the color temperature, then a mild CLAHE enhances the contrast without overexposing already bright areas. The changes are more subtle than in other weather conditions, but still effective. Objects that previously blended into the snowy background now appear more distinct and distinct. The 7.9-point confidence drop is explained by this altered color space; the model trained on dehazed features is less certain, despite the detection count doubling.

Why did bicycle performance collapse while motorcycle performance improved? Bicycles have thin, elongated structures. Frame, wheels, handlebars. Low contrast against backgrounds. In hazy conditions, these thin edges blur into noise. CLAHE and DCP sharpen edges, but if the sharpening introduces artifacts or merges bicycle parts with background textures, the model loses the holistic shape cue. Bicycles

appear at various orientations and scales, making them sensitive to edge perturbations.



Figure 9. Dehazing example for fog conditions



Figure 10. Dehazing example for rain conditions

### Weather-Adaptive Dehazing - Sand

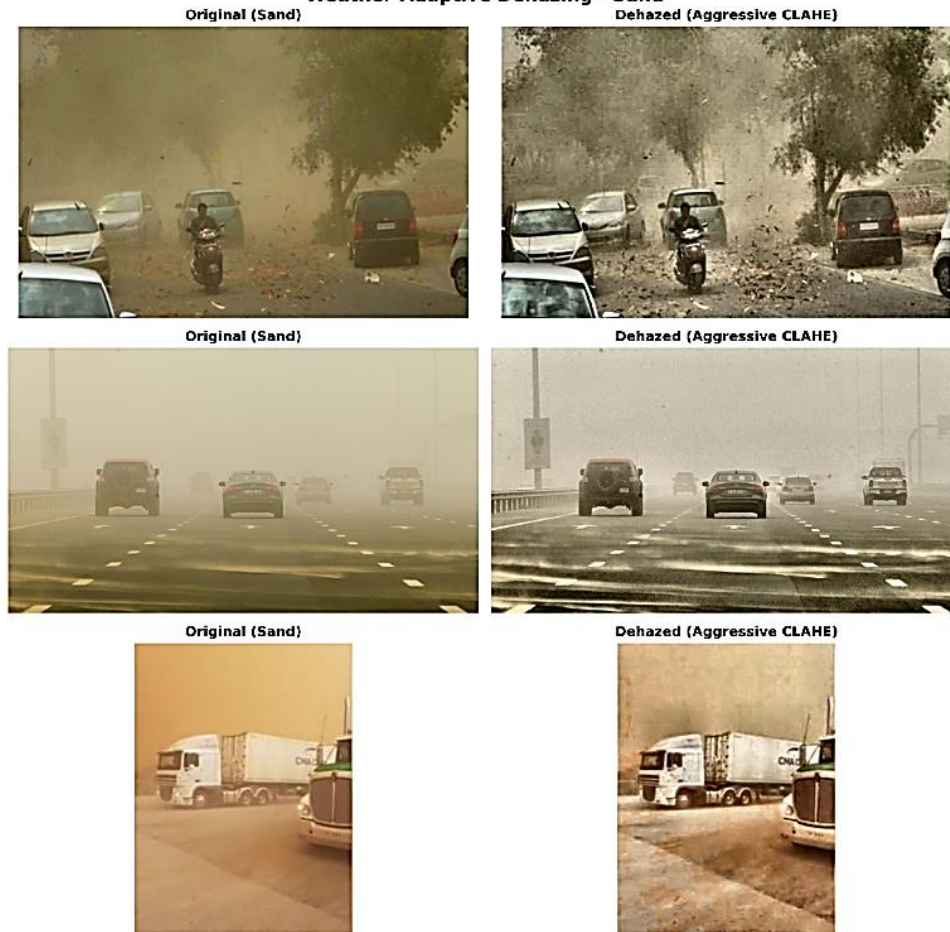
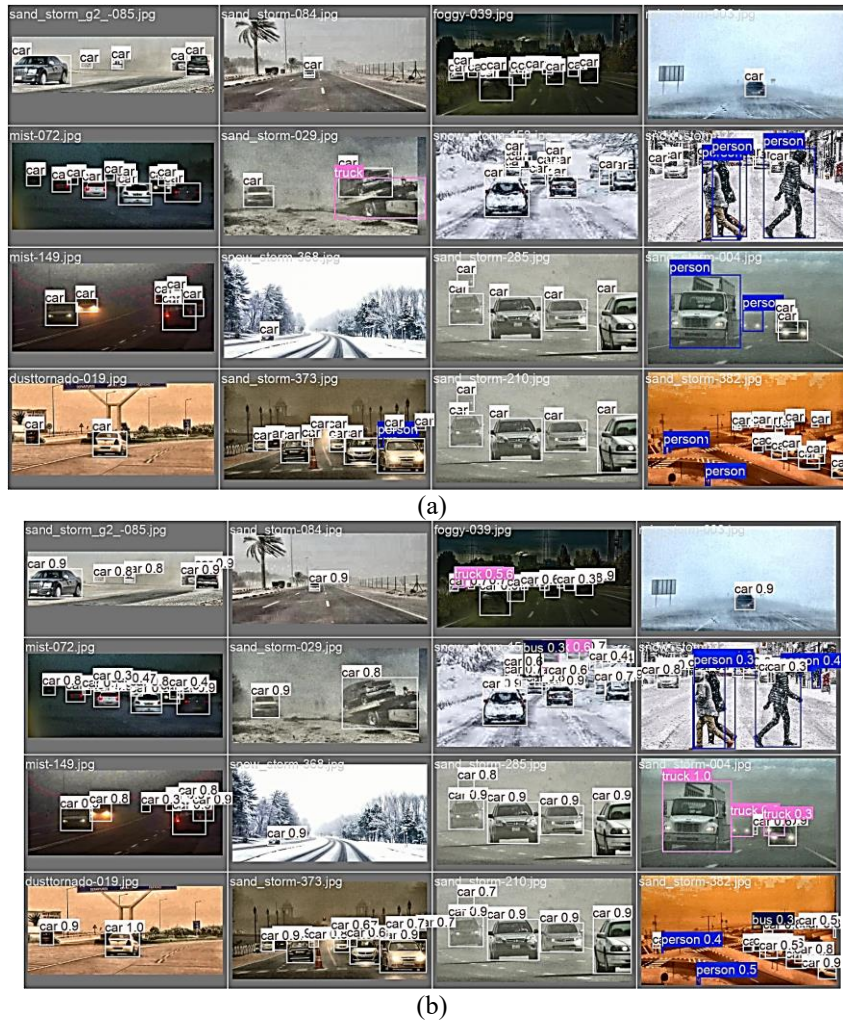


Figure 11. Dehazing example for sandstorm conditions

### Weather-Adaptive Dehazing - Snow



Figure 12. Dehazing example for snow conditions



**Figure 13.** Example of validation prediction where (a) label ground truth and (b) proposed mode prediction

Motorcycles are denser, with bulkier bodies and distinct engine components. Exhaust systems. These high-frequency details remain robust under dehazing. When haze is removed, the engine block and seat become more salient, improving detectability. Motorcycles are less prone to background confusion because their silhouette is more compact.

Cars remain the best-detected class at 88.6 percent even with dehazing. Large size, well-defined bounding boxes, and abundance in the training set contribute. 697 instances in baseline test set. Dehazing provides minimal benefit since cars are usually visible even in adverse weather. The small 3-point drop suggests marginal disruption from altered features.

The confusion matrix confirms bicycle's catastrophic failure. Of approximately 60 bicycle instances in the test set, only 2 were correctly detected with dehazing. The rest were either missed entirely, with no detection box generated, or misclassified as background because of IoU failures. Motorcycles succeeded because their denser, bulkier bodies remained robust. Engine block, seat, fuel tank. These components become more salient when haze is removed.

The catastrophic bicycle failure versus motorcycle success provides strong empirical evidence for the thin-structure hypothesis, even without feature visualization. Future work with integrated Grad-CAM or feature map analysis would validate this mechanism. Full failure case visualization with image triplets showing original, dehazed, and detection overlay would strengthen the analysis. Analyzing the change

in feature map responses for bicycle objects before and after dehazing would reveal which channels or spatial regions cause the collapse.

Based on Figure 13, the detection quality patterns are clearly visible. Labels show ground truth annotations, predictions display model output with confidence scores. Some detections are precise with high IoU, others are shifted or oversized. Bicycles were often undetected or experienced significant localization errors. Car detection was generally accurate, although the bounding box position was sometimes not entirely accurate. For motorcycles, recall improved, but occasional localization errors still occurred. This pattern is consistent with the metrics reported in Section 4.

Person false negatives increased from 4 baseline to 15. Enhanced visibility paradoxically makes the model miss obvious targets in certain cases. Bus detection experienced a moderate decline. Large vehicles, which have less prominent visual characteristics than cars, tend to struggle when feature distributions change. Class imbalance in the DAWN dataset magnifies variation in less frequently occurring classes, and dehazing artifacts have a greater impact on underrepresented categories.

## 6.4 Comparison with learning-based dehazing methods

### 6.4.1 Classical vs. Deep learning approaches

Our study employs classical dehazing methods (DCP,

CLAHE variants) rather than modern deep learning approaches like FFA-Net or AOD-Net. This choice warrants explanation, particularly given the superior restoration quality demonstrated by learning-based methods.

AOD-Net [33] takes a different approach by reformulating the atmospheric scattering model. Instead of estimating transmission  $t(x)$  and atmospheric light  $A$  separately, AOD-Net unifies these parameters into a single variable  $K(x)$  and directly minimizes reconstruction errors in the pixel domain. This end-to-end formulation achieves 21.54 dB PSNR on synthetic images and importantly can be embedded within object detection pipelines like Faster R-CNN. When jointly optimized with detection tasks, AOD-Net simultaneously improves both dehazing quality and detection performance.

FFA-Net [34] represents the state-of-the-art in learned dehazing. Its architecture combines channel attention and pixel attention mechanisms through a feature fusion structure, achieving 36.39 dB PSNR on the SOTS indoor test dataset substantially exceeding classical methods like DCP (18.98 dB). The feature attention module treats different channels and pixels unequally, providing flexibility for handling uneven haze distributions. Local residual learning allows thin haze regions to bypass through multiple connections while the network focuses on thick haze areas. This adaptive weighting produces superior restoration, especially in regions with dense haze and rich texture detail.

#### 6.4.2 Why classical methods for this investigation

We deliberately chose classical methods for several reasons. First, our research aims to investigate fundamental trade-offs between detection completeness and localization accuracy. Classical methods offer interpretability that deep models lack. When DCP flips the atmospheric scattering equation, we can trace exactly how it alters intensity distributions and enhances high-frequency features. When CLAHE performs local histogram equalization, we understand precisely how it merges or splits adjacent edges. This transparency allows us to connect observed detection failures (e.g., bicycle collapse) to specific preprocessing artifacts (e.g., edge perturbations from aggressive enhancement).

Deep learning dehazing models operate as black boxes. FFA-Net's attention mechanisms adaptively weight features through learned parameters. We cannot easily isolate which architectural components cause which detection effects. If bicycle detection failed with FFA-Net preprocessing, was it the channel attention, pixel attention, or feature fusion? The investigation becomes confounded by model complexity.

Second, classical methods provide weather-specific customization without requiring weather-labeled training data. DCP works well for fog because fog follows atmospheric scattering physics. CLAHE variants handle rain, sand, and snow through targeted color corrections (yellow cast removal, white balance, contrast enhancement). Deep models like FFA-Net are typically trained on synthetic hazy images from indoor/outdoor datasets that may not capture the full diversity of real-world weather degradations. AOD-Net's training uses NYU2 depth data with synthetic haze parameters ( $\beta \in \{0.4, 0.6, \dots, 1.6\}$ ), which approximates uniform fog but not rain droplets, sandstorm particles, or snow glare.

Third, computational efficiency matters for practical deployment. Classical methods run in milliseconds per frame. DCP processes a  $640 \times 640$  image in  $\sim 50$ -100 ms on CPU. CLAHE runs even faster. AOD-Net achieves 0.026 seconds per  $480 \times 640$  image on GPU, which is impressive for a deep

model but still requires GPU infrastructure. FFA-Net, with its multi-group architecture ( $3 \text{ groups} \times 19 \text{ blocks}$ ), demands more resources. For embedded systems in autonomous vehicles, classical preprocessing may be more viable.

#### 6.4.3 Complementary nature and future directions

Our findings with classical methods reveal that dehazing improves recall and precision but degrades mAP because of localization errors. This trade-off likely persists with learning-based dehazing, though the magnitude may differ. FFA-Net's superior restoration quality at 36.39 dB versus DCP's 18.98 dB suggests it preserves edges and textures better, potentially reducing bicycle-type failures. Yet even perfect dehazing creates a feature space mismatch. Detection models trained on original images will struggle with dehazed images unless retrained.

AOD-Net [33] offers an intriguing solution through joint optimization. When AOD-Net is embedded with Faster R-CNN and trained end-to-end, both dehazing and detection improve simultaneously. The dehazing module learns to improve features that specifically benefit detection rather than maximizing PSNR or SSIM. This is the logical next step. Combine weather-adaptive preprocessing with joint training.

Future work should integrate classical weather-specific methods with deep architectures. A hybrid pipeline could process fog images through DCP preprocessing, then feed into AOD-Net for refinement, finally entering YOLOv9 with joint optimization. Rain images get CLAHE enhancement before AOD-Net. Sand images receive yellow cast correction. Each weather type follows a customized path before unified deep learning. This combines the interpretability and weather specificity of classical methods with the representational power of deep learning.

Another direction involves architectural modifications informed by our findings. We discovered that thin structures like bicycles fail catastrophically while dense objects like motorcycles benefit. Could we design attention mechanisms that specifically preserve thin edges? FFA-Net's [34] pixel attention focuses on thick haze regions and high-frequency areas. A bicycle-aware attention module might prioritize elongated, low-contrast structures. SE attention used by Ning Kang et al. provides channel-wise feature recalibration [18]. Combining SE with FFA-Net's feature fusion could yield an architecture robust to both weather degradation and structural diversity.

The comparison with Kang et al. [18] suggests that weather-adaptive dehazing and architectural improvements address different failure modes. Dehazing improves visibility, increasing recall by 26.3 points relative to their baseline. SE attention improves feature weighting, boosting precision. Together, they might achieve both high recall and high mAP, resolving the trade-off we observed.

### 6.5 Positioning this work

Our contribution lies not in proposing the best dehazing method but in systematically revealing how dehazing affects detection metrics. Classical methods serve as controlled interventions. By using interpretable algorithms with known properties, we isolate the causal mechanisms behind the precision-recall-mAP divergence. FFA-Net and AOD-Net could replace our classical preprocessing in future experiments, but the core insight remains valid. Visibility improvement and spatial precision are orthogonal.

This work complements the deep learning literature. Where FFA-Net and AOD-Net demonstrate what performance is achievable, we explain why certain trade-offs occur and under which conditions dehazing helps versus harms. Practitioners can use our weather-specific and class-specific findings to decide when to apply dehazing. Small dense objects in snow or sand benefit. Thin structures in fog suffer. Researchers can build on our analysis to design better attention mechanisms or joint optimization strategies.

Classical and deep learning dehazing methods are complementary rather than competitive. Classical methods provide interpretability and weather specificity. Deep learning methods provide superior restoration and end-to-end optimization. Combining both approaches represents the most promising path forward. Using classical methods to inform deep architectures and employing joint training to align preprocessing with downstream tasks.

## 7. CONCLUSIONS

This study examines how weather-adjusted dehazing affects YOLOv9's ability to recognize objects in adverse weather conditions. We used the DAWN dataset, which includes four types of atmospheric degradation: fog, rain, sand, and snow. Stratified data partitioning ensured that each weather condition was proportionally represented. Different dehazing techniques were applied for each condition: DCP for fog combined with CLAHE and color correction; more aggressive CLAHE with color balancing was used for rain; soft CLAHE white balance was applied for sand; and light CLAHE white balance was used for snow.

The baseline model without dehazing achieved an mAP@0.5 of 69.4%, 21 points higher than previous results—this was made possible by a more optimized method. The YOLOv9-C architecture, stratified data partitioning, and hyperparameter optimization over 100 epochs resulted in the best convergence. Key findings revealed significant trade-offs. Recall increased by 12.3% and precision by 7.1%, but mAP decreased by 14.7%. The model found more objects and produced fewer false positives, but its ability to maintain detection quality decreased. The IoU threshold demands spatial precision, which does not always equate to improved object discrimination.

Class-specific patterns of effects provide additional insights. Dehazing improved performance for motorcycles, with an increase of 3.3 points in AP@0.5 and 6.0 points under the more stringent criteria. The dense, high-frequency structure of motorcycles remained stable after processing. In contrast, bicycle performance dropped significantly, losing 49.4 points, as edge artifacts caused by dehazing more easily damage thin, elongated structures. Cars continued to dominate, despite a slight decrease, while buses and large vehicles showed moderate degradation. Object size and structural complexity significantly determined the impact.

The impact depending on weather conditions also showed significant differences. The number of detections in snow nearly doubled (+ 127), but the confidence level dropped by 7.9 points. Visibility improved, but the color space change made the model more uncertain. Rain and sand showed a comparable increase in detections, despite a slight decrease in confidence. Fog became an anomaly: the number of detections dropped by 69, but the confidence level remained at 0.757. An overly aggressive DCP can remove real objects while

simultaneously reducing airborne scattering.

The practical implications are clear. In snowy or sandy conditions, where improved visibility is more important than impaired localization accuracy, dehazing is more suitable for small, dense objects. Conversely, avoid thin objects or foggy conditions, as conventional methods can produce detrimental artifacts. The most important point is that the pre-processing and evaluation stages must be aligned. If a model is trained with pre-processed features but tested using the original annotations, spatial mismatches will arise, ultimately harming metrics that require localization accuracy.

Going forward, integrated approaches that combine dehazing with architectural improvements such as SE attention or MPDIoU loss should be explored. Such network modifications could potentially align with the adaptive preprocessing used. A combined approach that simultaneously considers input quality, feature weighting, and bounding box regression has the potential to provide mutually reinforcing benefits.

## 8. FUTURE WORKS

Fine-tuning on dehazed datasets offers a path forward. Freeze the backbone. Train only detection heads, bounding box regression plus classification, on dehazed images for 20-30 epochs. This transfer learning approach adapts localization to modified feature space without losing the baseline's strong performance. Expected outcome: recover lost mAP while maintaining recall and precision gains.

Optimization needs attention parameters. Current work uses fixed settings per weather type: DCP  $\omega = 0.95$ , CLAHE clip limits ranging from 2.0 to 4.0. Sweep these values. Vary DCP  $\omega$  from 0.85 to 0.98, CLAHE clip limits from 1.5 to 5.0. Evaluate the validation set to find optimal configurations. Better yet, implement per-image adaptive tuning based on local haze density or contrast measures. Dense fog images might need gentler DCP processing. Light haze could tolerate aggressive parameters. Validation-based optimization per weather type reduces artifacts that harm thin-structured objects.

Deep learning dehazing methods warrant comparison. AOD-Net, GridDehazeNet, FFA-Net learn restoration mappings from data rather than relying on hand-crafted priors. Train these networks on paired hazy-clear images, synthetic or from other datasets, then apply to DAWN. Hypothesis: learned dehazing preserves spatial information better due to end-to-end optimization. Classical methods are effective for specific degradation types yet lack flexibility. Neural networks can potentially avoid IoU degradation observed with DCP and CLAHE. Besides standalone comparison, integrated approach combining dehazing with SE attention plus MPDIoU loss could yield synergistic gains. Enhanced visibility meets feature recalibration meets improved localization loss. Three complementary strategies addressing different aspects of the detection pipeline.

## REFERENCES

- [1] Chen, Q., Song, Z., Dong, J., Huang, Z., Hua, Y., Yan, S. (2014). Contextualizing object detection and classification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 37(1): 13-27.

- <https://doi.org/10.1109/TPAMI.2014.2343217>
- [2] Zhang, H., Xiao, L., Cao, X., Foroosh, H. (2022). Multiple adverse weather conditions adaptation for object detection via causal intervention. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 46(3): 1742-1756. <https://doi.org/10.1109/TPAMI.2022.3166765>
- [3] Sahu, G., Seal, A., Bhattacharjee, D., Frischer, R., Krejcar, O. (2022). A novel parameter adaptive dual channel MSPCNN based single image dehazing for intelligent transportation systems. *IEEE Transactions on Intelligent Transportation Systems*, 24(3): 3027-3047. <https://doi.org/10.1109/TITS.2022.3225797>
- [4] Li, J., Xu, R., Liu, X., Ma, J., et al. (2024). Domain adaptation based object detection for autonomous driving in foggy and rainy weather. *IEEE Transactions on Intelligent Vehicles*, 10(2): 900-911. <https://doi.org/10.1109/TIV.2024.3419689>
- [5] Almalioglu, Y., Turan, M., Trigoni, N., Markham, A. (2022). Deep learning-based robust positioning for all-weather autonomous driving. *Nature Machine Intelligence*, 4(9): 749-760. <https://doi.org/10.1038/s42256-022-00520-5>
- [6] Zhang, C., Wang, H., Cai, Y., Chen, L., Li, Y. (2024). TransFusion: multi-modal robust fusion for 3D object detection in foggy weather based on spatial vision transformer. *IEEE Transactions on Intelligent Transportation Systems*, 25(9): 10652-10666. <https://doi.org/10.1109/TITS.2024.3420432>
- [7] He, K., Sun, J., Tang, X. (2010). Single image haze removal using dark channel prior. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(12): 2341-2353. <https://doi.org/10.1109/TPAMI.2010.168>
- [8] Li, C., Zhou, H., Liu, Y., Yang, C., Xie, Y., Li, Z., Zhu, L. (2023). Detection-friendly dehazing: Object detection in real-world hazy scenes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(7): 8284-8295. <https://doi.org/10.1109/TPAMI.2023.3234976>
- [9] Wang, X., Liu, X., Yang, H., Wang, Z., et al. (2024). Degradation modeling for restoration-enhanced object detection in adverse weather scenes. *IEEE Transactions on Intelligent Vehicles*, 10(3): 2064-2079. <https://doi.org/10.1109/TIV.2024.3442924>
- [10] Kan, S., Zhang, Y., Zhang, F., and Cen, Y. (2022). A GAN-based input-size flexibility model for single image dehazing. *Signal Processing: Image Communication*, 102: 116599. <https://doi.org/10.1016/j.image.2021.116599>
- [11] Chen, Y., Li, W., Sakaridis, C., Dai, D., Van Gool, L. (2018). Domain adaptive faster R-CNN for object detection in the wild. In 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, pp. 3339-3348. <https://doi.org/10.1109/CVPR.2018.00352>
- [12] Liu, Z., Zhang, J., Zhang, X., and Song, H. (2026). Robust object detection in adverse weather conditions: ECL-YOLOv11 for automotive vision systems. *Sensors*, 26(1): 304. <https://doi.org/10.3390/s26010304>
- [13] Fang, W., Zhang, G., Zheng, Y., Chen, Y. (2023). Multi-task learning for UAV aerial object detection in foggy weather condition. *Remote Sensing*, 15(18): 4617. <https://doi.org/10.3390/rs15184617>
- [14] Liu, B., Jin, J., Zhang, Y., Sun, C. (2025). WRRT-DETR: Weather-robust RT-DETR for drone-view object detection in adverse weather. *Drones*, 9(5): 369. <https://doi.org/10.3390/drones9050369>
- [15] Yang, W., Tan, R.T., Feng, J., Liu, J., Guo, Z., Yan, S. (2017). Deep joint rain detection and removal from a single image. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1357-1366.
- [16] Gupta, H., Kotlyar, O., Andreasson, H., Lilienthal, A.J. (2024). Robust object detection in challenging weather conditions. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), Waikoloa, HI, USA, pp. 7523-7532.
- [17] Zhang, Z., Gong, H., Feng, Y., Chu, Z., Liu, H. (2024). Enhancing object detection in adverse weather conditions through entropy and guided multimodal fusion. In 17th Asian Conference on Computer Vision, Hanoi, Vietnam, pp. 22-38. [https://dl.acm.org/doi/10.1007/978-981-96-0972-7\\_2](https://dl.acm.org/doi/10.1007/978-981-96-0972-7_2)
- [18] Kang, N., Ma, F., Wan, W., Wang, D., Yao, H., Sheng, K. (2024). Improved YOLOv9-based objects detection in adverse weather conditions for autonomous driving. *IFAC-PapersOnLine*, 58(29): 267-271. <https://doi.org/10.1016/j.ifacol.2024.11.155>
- [19] Kumar, M., Yadav, A.L., Arora, A., Deb, A. (2023). Object detection in adverse weather conditions using machine learning. In 2023 International Conference on Integrated Intelligence and Communication Systems (ICIICS), Kalaburagi, India, pp. 1-7. <https://ieeexplore.ieee.org/document/10421471>
- [20] Chu, Z. (2024). D-YOLO a robust framework for object detection in adverse weather conditions. arXiv preprint arXiv:2403.09233. <https://doi.org/10.48550/arXiv.2403.09233>
- [21] Jing, Z., Li, S., Zhang, Q. (2024). YOLOv8-STE: Enhancing object detection performance under adverse weather conditions with deep learning. *Electronics*, 13(24): 5049. <https://doi.org/10.3390/electronics13245049>
- [22] He, K., Sun, J., Tang, X. (2012). Guided image filtering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(6): 1397-1409. <https://doi.org/10.1109/TPAMI.2012.213>
- [23] Tan, R.T. (2008). Visibility in bad weather from a single image. In 2008 IEEE Conference on Computer Vision and Pattern Recognition, Anchorage, AK, USA, pp. 1-8. <https://doi.org/10.1109/CVPR.2008.4587643>
- [24] Zuiderveld, K. (1994). Contrast limited adaptive histogram equalization. In Graphics Gems, pp. 474-485. <https://doi.org/10.1016/B978-0-12-336156-1.50061-6>
- [25] Pizer, S.M., Amburn, E.P., Austin, J.D., Cromartie, R., et al. (1987). Adaptive histogram equalization and its variations. *Computer Vision, Graphics, and Image Processing*, 39(3): 355-368. [https://doi.org/10.1016/S0734-189X\(87\)80186-X](https://doi.org/10.1016/S0734-189X(87)80186-X)
- [26] Jin, S.H., Son, D.M., Lee, S.H., Go, Y.H., Lee, S.H. (2025). Enhancing local contrast in low-light images: A multiscale model with adaptive redistribution of histogram excess. *Mathematics*, 13(20): 3282. <https://doi.org/10.3390/math13203282>
- [27] Buchsbaum, G. (1980). A spatial processor model for object colour perception. *Journal of the Franklin Institute*, 310(1): 1-26. [https://doi.org/10.1016/0016-0032\(80\)90058-7](https://doi.org/10.1016/0016-0032(80)90058-7)

- [28] Land, E.H., McCann, J.J. (1971). Lightness and retinex theory. *Journal of the Optical Society of America*, 61(1): 1-11. <https://doi.org/10.1364/JOSA.61.000001>
- [29] Wang, C.Y., Yeh, I.H., Liao, H.Y.M. (2024). YOLOv9: Learning what you want to learn using programmable gradient information. *arXiv preprint arXiv:2402.13616*. <https://doi.org/10.48550/arXiv.2402.13616>
- [30] Lin, T.Y., Goyal, P., Girshick, R., He, K., Dollár, P. (2017). Focal loss for dense object detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 42(2): 318-327. <https://doi.org/10.1109/TPAMI.2018.2858826>
- [31] Li, X., Lv, C., Wang, W., Li, G., Yang, L., Yang, J. (2022). Generalized focal loss: Towards efficient representation learning for dense object detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(3): 3139-3153. <https://doi.org/10.1109/TPAMI.2022.3180392>
- [32] Peng, H., Yu, S. (2021). A systematic IOU-related method: Beyond simplified regression for better localization. *IEEE Transactions on Image Processing*, 30: 5032-5044. <https://doi.org/10.1109/TIP.2021.3077144>
- [33] Li, B., Peng, X., Wang, Z., Xu, J., Feng, D. (2017). Aod-net: All-in-one dehazing network. In *Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy*, pp. 4770-4778. <https://doi.org/10.1109/ICCV.2017.511>
- [34] Qin, X., Wang, Z., Bai, Y., Xie, X., Jia, H. (2019). FFA-Net: Feature fusion attention network for single image dehazing. *arXiv preprint arXiv:1911.07559*. <https://doi.org/10.48550/arXiv.1911.07559>