Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

W-ASFF: A Weather-Adaptive Spatial Feature Fusion Network for Robust Object Detection in Adverse Driving Conditions

Syeda Nafeesa Batool Kazmi
Air University
arif ArifUllah
arifullahms88@gmail.com

Multimedia University
Siti Fatimah Abdul Razak
Multimedia University
Wael Anabousi
Multimedia University of Kenya
Sumendra Yogarayan
Multimedia University of Kenya
Abdullah ullah
Multimedia University

Article

Keywords: Object detection, adverse weather, autonomous driving, channel attention, spatial feature fusion

Posted Date: November 20th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11458905/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

W-ASFF: A Weather-Adaptive Spatial Feature Fusion Network for Robust Object Detection in Adverse Driving Conditions

Syeda Nafeesa Batool Kazmi[1], ArifUllah[2], Siti Fatimah Abdul Razak[2] Wael Anabousi[3], Sumendra Yogarayan[4], Abdullah[5]
[12#]Department of Computer Science, Faculty of Computing and Artificial Intelligence, Air University, Islamabad, Pakistan.
[234]#Centre for Intelligent Cloud Computing, Multimedia University, Melaka, 75450, Malaysia
[4#]Hourani Center for Applied Scientific Research, Al-Ahliyya Amman University, Amman 19328, Jordan
[5#]Department of Computer Science UTHM Malaysia
Email: fatimah.razak@mmu.edu.my, arifullahms88@gmail.com

Abstract: Most object detectors are trained and evaluated on clear-weather imagery, yet autonomous vehicles have to operate reliably in rain, fog, snow, and low light conditions where these models routinely struggle. We address this with W-ASFF, a weather-adaptive spatial feature fusion network built on a YOLO11s backbone. Instead of treating every scene identically, W-ASFF conditions its multi-scale feature fusion on the visual weather context. A lightweight Weather Embedding module first projects the scene's weather identity into a compact latent vector. A Weather Gate module then uses that embedding, together with global average pooling of the P3, P4, and P5 feature maps, to recalibrate channel responses through a learned sigmoid gate. The recalibrated features feed into a weighted Adaptive Spatial Feature Fusion (W-ASFF) stage, where per-scale fusion coefficients are learned via a softmax normalized scoring function that keeps them non-negative and summing to one. Training is end-to-end, combining CIoU localization loss, detection confidence and classification loss, and auxiliary weather and time-of-day classification losses. On the BDD100K dataset, W-ASFF reaches 54.7% mAP@0.5, 66.3% precision, 52.1% recall, and an F1-score of 0.58, sustaining approximately 228 FPS (4.39 ms inference, batch size 16) with 93.8% accuracy on the auxiliary weather-classification task.
Keywords: Object detection, adverse weather, autonomous driving, channel attention, spatial feature fusion.

INTRODUCTION

Safe autonomous driving depends on reliable perception, and object detection carries much of that load identifying vehicles, pedestrians, and other road users directly from onboard camera streams. Convolutional and transformer-based detectors now achieve strong accuracy on curated, largely clear weather benchmarks, and the YOLO family in particular has become a de facto standard for real-time deployment, pairing competitive accuracy with inference speeds suitable for embedded automotive hardware [1].Real driving conditions are far less forgiving: rain, fog, snow, and low illumination degrade contrast, blur object boundaries, and introduce sensor noise that pushes the input distribution away from what these detectors were actually trained on [2].This degradation is not specific to cameras a broad survey of adverse weather perception across LiDAR, radar, and camera modalities finds that explicit weather condition classification remains comparatively underexplored relative to sensor-hardware and fusion solutions [3]. A large body of work has addressed adverse-weather robustness by either restoring the image before detection dehazing, deraining, low, light enhancement or by augmenting training data with synthetic weather effects [4]. Both strategies treat weather as a nuisance to be removed rather than as informative context, and restoration based pipelines add a separate pre-processing stage that increases latency and can itself introduce artifacts [5]. A complementary line of research instead

modifies the detector itself through attention modules, multi-scale feature fusion, or weather-conditioned branches so the network adapts its internal feature computation to the scene rather than relying on a fixed processing pipeline [6].Adaptive Spatial Feature Fusion (ASFF) demonstrated that learning scale-wise fusion weights instead of fixed feature pyramid addition improves multi-scale detection and subsequent work has combined ASFF-style fusion with YOLO backbones for autonomous driving scenes [7].Existing weather-aware detectors mostly condition on weather implicitly through data augmentation or generic attention rather than an explicit, learned signal that directly shapes how multi-scale features are fused [8]. What's missing is a lightweight, camera-only mechanism that represents the scene's weather condition explicitly, uses that representation to recalibrate backbone features before fusion, and then fuses those recalibrated features adaptively, without a separate restoration network or an added sensor modality. That gap is what this work targets. We propose W-ASFF, a weather-adaptive spatial feature fusion network built on a YOLO11s backbone [9]. A Weather Embedding module encodes weather identity into a 64-dimensional latent vector; a WeatherGate module uses this, with global average pooling of the P3, P4, and P5 features, to produce channel-wise gates that recalibrate each scale before fusion. A weighted ASFF stage then learns per-scale fusion coefficients via a softmax-normalized scoring function, combining the gated features into a single representation for the detection head. Training uses a joint loss combining CIoU localization, detection confidence and classification terms, and auxiliary weather and time-of-day classification losses [10].

Related Work

YOLO-Based Detectors

Object detectors are typically divided into two camps. Two stage methods first propose candidate regions, and then classify and refine them, which tends to achieve strong localization accuracy but at the cost of latency which does not suit the real-time demands of automotive perception. The YOLO family belongs to the camp of single-stage detector, which gets rid of the proposal step and directly predicts classes and boxes from a dense grid in one pass and the reason why it becomes a near default choice for embedded deployment. YOLO has evolved through its iterations from anchor-based single-shot grid detector to anchor-free multi-task architectures that combine detection, segmentation and pose estimation in a single backbone [11]. YOLO11 follows this trend, improving the CSP based backbone and neck to reduce the number of parameters Without loss of accuracy, and it is available in sizes from nano to extra-large so practitioners can choose a point on the speed–accuracy curve best suited to their hardware [12]. Its cross-stage partial (CSP) blocks split an incoming feature map into two branches, run one through a stack of bottleneck layers, then concatenate it back with the un-touched branch; this keeps gradient flow rich while reducing redundant computation across layers [13]. YOLO11s is the backbone here as it provides a good trade-off between accuracy and real-time throughput. Its P3, P4 and P5 outputs are fed directly to the weather-conditioned fusion stage described in Section.

Weather-Robust Object Detection

Several recent efforts adapt YOLO to the specific failure mode addressed in this paper: performance drop-off once weather departs from the clear-sky conditions on which most detectors are trained. YOLOv11-TWCS is based on CBAM and a decoupled down sampling block with a Trans Weather-like attention module, with improvements under rain, fog and low light on DAWN, KITTI and Udacity [14]. Conversely, ECL-YOLOv11 adds edge-enhancement and context-guided modules to recover the boundary information that weather-induced blur tends to wash out, as the

rationale is that sharper edges still help localization even when over-all contrast is poor [15]. Another approach frames the problem as one of restoration rather than architecture: IA-YOLO learns differentiable image-processing filters, with a small CNN predicting the filter's parameters from the input's brightness, tone, and weather cues, and trains this jointly with the detector so that the enhancement step is optimized for detection rather than for looking good to a human viewer [16].Diffusion-based augmentation is a data-side alternative to the same idea, synthesizing adverse-weather training images from clear-weather sources so that a standard detector sees a training distribution closer to what it will actually encounter, without needing real annotated bad-weather data [17]. The common thread in these methods is the belief that weather needs to be eliminated, either by bringing the image back to its clear-weather appearance, or by demonstrating the network synthetic weather until the latter loses its relevance. In contrast to these, W-ASFF does not attempt to restore or render the input in any way. The weather conditions are viewed as information affecting the combination of the multi-scale features [18].

Attention Mechanisms and Feature Fusion

Channel and spatial attention modules re-calibrate feature map responses via condensing global context into a compact descriptor and running it through a small gating network, typically a bottleneck consisting of fully-connected or pointwise convolutional layers followed by a sigmoid output layer that serves to re-weight the original channel responses. The Squeeze-and-Excitation mechanism pioneered this design, and CBAM builds on it via adding a second attention head following the first one, thus learning not only which channels to pay attention to, but also where in the image those channels should be paid attention to [19]. The WeatherGate architecture employs a similar squeeze-excite structure, although the gating signals are not derived solely from the statistics of the feature maps them-selves, but are also conditioned on the learned weather embedding. With regard to fusion methods, the Feature Pyramid Network established the archetypal model followed by most detectors, where course and semantic rich features from deep layers are up sampled and fused with fine and accurate spatial information from shallow layers via lateral connections, thus having both types of information at each pyramid level [20]. The issue here is that fusion is achieved by means of simple addition, treating all scales equally without taking into consideration actual contents of the image. The adaptive spatial feature fusion (ASFF) solves this problem through per-scale weighting, resizes each pyramid level to a uniform resolution, then uses a tiny network to predict a softmax-normalized weight for each pyramid at each spatial location, resulting in a weighted combination of the input pyramids. ASFF consistently outperforms fixed fusion in detecting objects that fall between scale levels [21]. In more recent works, ASFF-based fusion has been incorporated into the YOLO backbones dedicated for driving scenarios: one such variant that uses YOLOv8 backbone with attention guided pyramid fusion along with dynamic label assign-mint for crowded driving scenarios achieves 71.2% mAP on BDD100K [22], while YOLO-Vehicle-Pro couples debasing with multi-modal fusion for cloud-edge deployments in low visibility conditions [23]. The work most similar to ours in architecture is MAAFFNet which incorporates ASFF fusion module into YOLOv8 backbone along with the iterative fusion with attention and loss function designed to emphasize hard samples, achieving significantly improved mAP over its baseline on KITTI, COCO-traffic and PASCAL VOC datasets – but not once throughout all the experiments weather condition of the data gets mentioned [24]. Thus, although the body of work continues to make improvements in learning the fusion weights from feature contents – shapes, scales, local context – the weather conditions under which the scene is captured are still neglected. This is precisely the problem that W-ASFF is designed to

solve, as the fusion coefficients are jointly learned with a weather condition gate so that the features in the backbone can be calibrated for weather conditions [25].

Multi-Scale Detection on Driving Datasets

The two benchmarks that driving-scene detectors are typically evaluated against are BDD100K and KITTI, which, however, measure very different capabilities. While KITTI contains only samples gathered in mostly clear, well-lit conditions in one geographical region, with relatively low variability in scene types, having high score on KITTI indicates little about how robust the detector will perform in poor weather conditions. The reason why BDD100K stands out as a better benchmark for the task tackled in this paper is that it was gathered with a much larger variety in weather, time of day, and scene types, and with weather and time-of-day labels provided [26]. As far as reported mAP of BDD100K goes, it varies between 58% and 71% depending on the backbone, resolution and class grouping [27], [28], and these numbers may even slightly overestimate the actual performance because of the positive effect of the vast majority of clear-weather daytime samples on aggregate mAP score. This contrast between general difficulty of the BDD100K benchmark and adverse-weather difficulty of BDD100K in particular is what motivates our decision to use the BDD100K dataset to evaluate W-ASFF, since its full range of weather, time-of-day, and scene diversity provides a realistic basis for assessing weather-adaptive detection performance [29].

PROPOSED METHODOLOGY

The W-ASFF framework extracts multi-scale feature maps, namely P3, P4, and P5, having 640×640 inputs, using a YOLO11s CSPDarknet backbone, with strides being 8, 16, and 32, respectively. Simultaneously, there is another Weather Embedding framework which captures the weather identity of the scene in a compact form. This identity is then used to condition a WeatherGate block that recalibrates the channels of the multi-scale features using the squeeze and excite style transformation. The recalibrated multi-scale features are then fused using the weighted ASFF (W-ASFF) framework using softmax normalized fusion coefficients [30].

A. Backbone Feature Extraction

The input image $I \in \mathbb{R}^{640\times640\times3}$ is processed by a CSPDarknet backbone with a stride convolutional stem followed by four CSP stages. Each convolutional layer computes a standard 2-D correlation between the input feature map X and a learned kernel bank K, followed by a bias and non-linearity:

$$Y(i,j,o) = \varphi\left(\sum_{m,n,c} K\,(m,n,c,o)\cdot X(i\cdot s+m,\ j\cdot s+\ \ n,\ c) + b(o)\right) \quad (1)$$

Where s is the stride, $\varphi(\cdot)$ is the SiLU activation used throughout the backbone, and o indexes output channels [31].Figure 1present the proposed model structure.

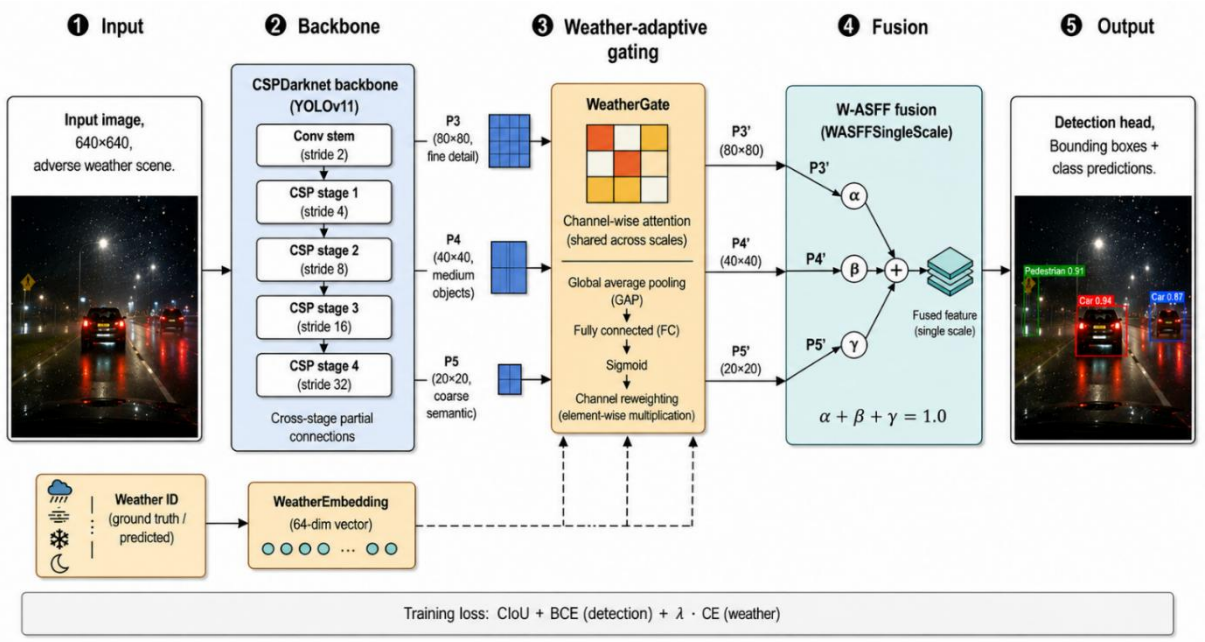


Fig. 1. Overall architecture of the proposed W-ASFF network: input, CSPDarknet backbone (P3/P4/P5 extraction), Weather Embedding, WeatherGate channel recalibration, W-ASFF fusion, and the anchor-free detection head

.

Each CSP stage splits its input channels into two partitions, passes one partition through a sequence of such bottleneck convolutions, and concatenates the result with the unprocessed partition to reduce redundant gradient flow:

$$X_{out} = \text{Concat}\big(X_a,\ T(X_b)\big) \tag{2}$$

Where X_a and X_b are channel-wise partitions of the stage input and $T(\cdot)$ denotes the composition of bottleneck convolutions in Eq. (1). Three feature maps are tapped from stages of stride 8, 16, and 32:

$$P_l = f_{bb}^{l}(I) \in \mathbb{R}^{H_l \times W_l \times C_l}, l\{3,4,5\} \tag{3}$$

Where f_{bb}^{l} denotes the backbone computation up to the tap point for scale$l$, and H_l, W_l, C_l are the height, width, and channel count of that scale. P3 (80×80) retains fine spatial detail suited to small objects, P4 (40×40) captures medium-scale structure, and P5 (20×20) encodes coarse semantic context [32].

Weather Embedding

The weather identity signal, available as ground-truth weather category during training, is represented as a one-hot vector over K weather classes:

$$\mathbf{w} = [w_1, \dots, w_K]^{\top}\{0,1\}^{K}, \quad \sum_{k=1}^{K} w_k\, 1 \tag{4}$$

And linearly projected into a 64-dimensional latent embedding:

$$e_w = W_e\, w + b_e \tag{5}$$

where $W_e \in \mathbb{R}^{64\times K}$ and $\mathbf{b}_e \in \mathbb{R}^{64}$ are learned parameters and $\mathbf{e}_w \in \mathbb{R}^{64}$ is the resulting weather embedding fed to every WeatherGate instance [33].

WeatherGate: Weather-Conditioned Channel Recalibration

For each scale $l \in \{3,4,5\}$, the WeatherGate module first computes a per-channel descriptor via global average pooling (GAP) of the backbone feature map:

$$\mathbf{z}_l = \frac{1}{H_l W_l} \sum_{i=1}^{H_l} \sum_{j=1}^{W_l} P_l\,(i,j,:) \tag{6}$$

where$\mathbf{z}_l \in \mathbb{R}^{C_l}$. The pooled descriptor is concatenated with the weather embedding to form the gate input:

$$\tilde{\mathbf{z}}_l = [\mathbf{z}_l\,;\, \mathbf{e}_w] \;\in \mathbb{R}^{C_l+64} \tag{7}$$

And passed through a shared bottleneck of reduction ratio r, with ReLU defined as max(0,x):

$$\mathbf{h}_l = \max(0,\; W_1 \tilde{\mathbf{z}}_l + \mathbf{b}_1), \quad W_1 \in \mathbb{R}^{(C_l/r)\times(C_l+64)} \tag{8}$$

Followed by a second fully-connected layer whose output is passed through a logistic sigmoid σ(x) = 1/(1+e−x) to produce a gate bounded in (0,1) per channel:

$$\mathbf{g}_l = \sigma(W_2 \mathbf{h}_l + \mathbf{b}_2) \in (0,1)^{C_l} \tag{9}$$

Where W1, W2, b1, b2 are shared across P3, P4, and P5 — the same FC–ReLU–FC–sigmoid weights are applied independently to each scale, so no additional per-scale parameters are introduced. The gate recalibrates the feature map through channel-wise (broadcast) multiplication:

$$P'_l(i,j,c) = P_l(i,j,c) \cdot g_l(c) \tag{10}$$

This recalibration suppresses channels that are uninformative under the detected weather condition and amplifies channels that carry weather-relevant discriminative signal [34]. Fig. 2 details the internal structure of this module.

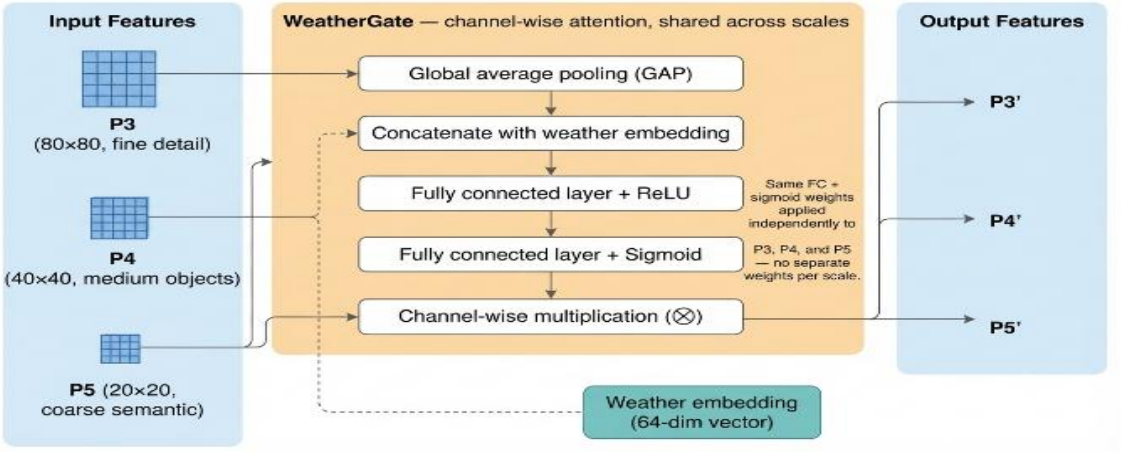
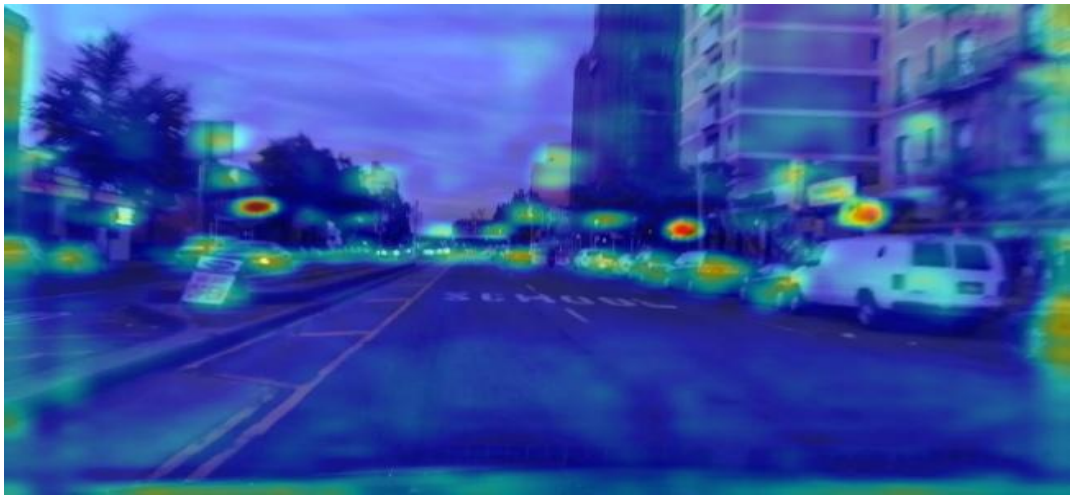


Fig. 2. Internal structure of the WeatherGate module

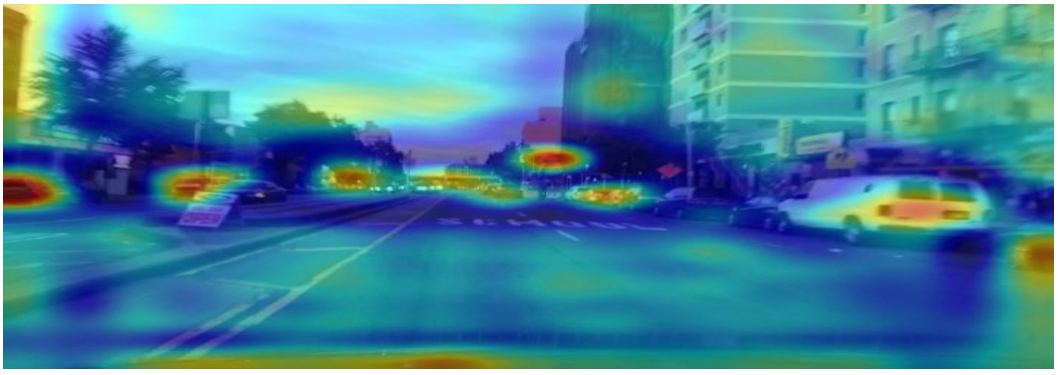
The raw multi-scale activations prior to this recalibration and fusion are visualized in Fig. 3 for a representative adverse-weather scene.



(a)



(b)



(c)

Fig. 3. Visualization of raw backbone feature activations before WeatherGate recalibration: (a) P3, (b) P4, (c) P5.

Weighted Adaptive Spatial Feature Fusion (W-ASFF)

After recalibration, the three feature maps P'_3, P'_4, P'_5 are resized to a common reference resolution (that of P3) using bilinear interpolation, denoted R(·). For a target location (x,y) in the resized map, bilinear interpolation combines the four nearest source samples weighted by their proximity:

$$R(P'_l)(x,y) = \sum_{a\in\{0,1\}} \sum_{b\in\{0,1\}} w_{a,b} \cdot P'_l(\lfloor x \rfloor + a,\ \lfloor y \rfloor + b) \tag{11}$$

where $w_{a,b} = (1 - |x - \lfloor x \rfloor - a|)(1 - |y - \lfloor y \rfloor - b|)$ are the standard bilinear interpolation weights. Each resized map is then combined using learned, normalized fusion coefficients. A lightweight scoring function first reduces each recalibrated, resized map to a scalar logit per scale via global average pooling followed by a learned projection:

$$u_l = \mathbf{w}_l^{\top} \cdot \mathrm{GAP}\big(R(P'_l)\big) + c_l, \quad l \in \{3,4,5\} \tag{12}$$

and the three logits are normalized with a softmax so that the resulting fusion weights are strictly non-negative and sum to one:

$$\alpha = \frac{e^{u_3}}{e^{u_3}+e^{u_4}+e^{u_5}}, \quad \beta \frac{e^{u_4}}{e^{u_3}+e^{u_4}+e^{u_5}} \gamma = \frac{e^{u_5}}{e^{u_3}+e^{u_4}+e^{u_5}}, \quad \alpha + \beta + \gamma = 1 \tag{13}$$

The fused feature map used by the detection head is the resulting convex combination: F_{fused} = $\alpha \cdot R(P'_3) + \beta \cdot R(P'_4) + \gamma \cdot R(P'_5)$ (14)

This formulation lets the network learn, for each scene, how much to trust fine-detail, medium-scale, and coarse-semantic information, which is particularly relevant when weather degrades different scales unevenly for instance, fog attenuates fine P3 detail more strongly than coarse P5 semantics [35]. The resulting fused representation is visualized in Fig. 4.

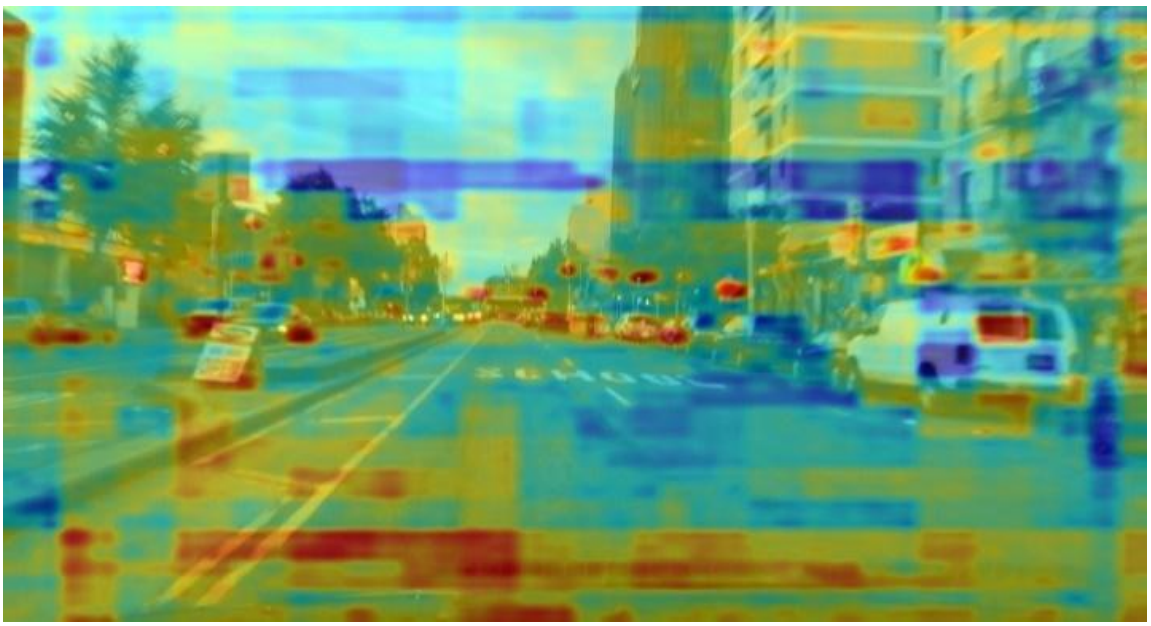


Fig. 4. Visualization of the fused feature

Detection Head

An anchor-free detection head that predicts, for each spatial cell, a bounding-box offset, an objectless confidence, and class probabilities, following the YOLO decoding convention, decodes the fused feature. The predicted box center is decoded relative to the cell origin (c_x, c_y) using a sigmoid-bounded offset:

$$b_x = \sigma(t_x) + c_x,\ b_y = \sigma(t_y) + c_y \tag{15}$$

and the predicted box width and height are decoded relative to a prior (p_w, p_h) through an exponential mapping that guarantees positivity:

$$b_w = p_w\, e^{t_w}, \qquad b_h = p_h\, e^{t_h} \tag{16}$$

Objectless confidence and per-class probabilities are each produced by an independent sigmoid, allowing multi-label class prediction rather than a single softmax over mutually exclusive classes:

$$\Pr(\text{obj}) = \sigma(t_o) \tag{17}$$

$$\Pr(c_k \mid \text{obj}) = \sigma(t_{c_k}), k1, \ldots, 10 \quad [36]. \tag{18}$$

Fig. 5 shows a qualitative example of the decoded output for an adverse-weather scene.

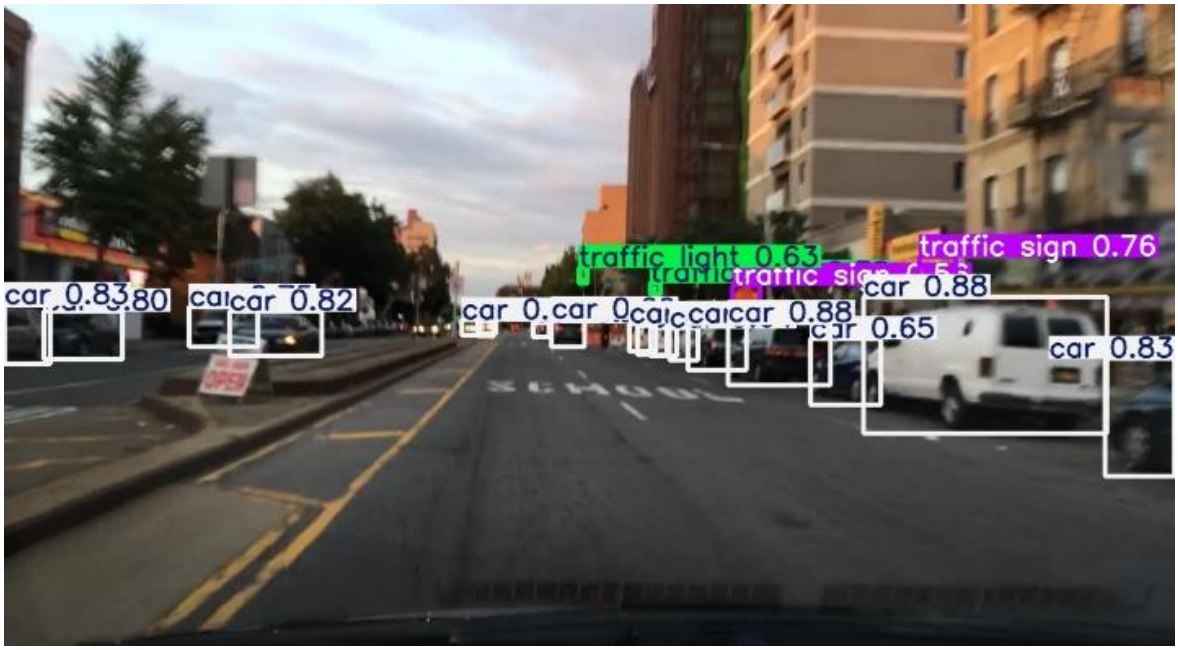


Fig. 5. Qualitative detection output of W-ASFF on an adverse-weather scene

Training Objective

Localization is supervised with Complete-IoU (CIoU) loss, which builds on the base Intersection-

over-Union between predicted box b and ground-truth box b^{gt}:

$$\text{IoU} = \frac{|b \cap b^{gt}|}{|b \cup b^{gt}|} \tag{19}$$

CIoU augments IoU with a normalized central-point distance penalty and an aspect-ratio consistency term:

$$\mathcal{L}_{CIoU} = 1 - \text{IoU} + \frac{\rho^2(b, b^{gt})}{c^2} + \alpha_v\, v \tag{20}$$

Where $\rho(b, b^{gt})$ is the Euclidean distance between predicted and ground-truth box centers, c is the diagonal length of the smallest enclosing box covering both b and bgt, and the aspect-ratio consistency term v penalizes mismatched width-to-height ratios:

$$v = \frac{4}{\pi^2}\left(\arctan\frac{w^{gt}}{h^{gt}} - \arctan\frac{w}{h}\right)^2 \tag{21}$$

And its trade-off weight α_v is defined so that it vanishes for well-overlapping boxes and grows as the aspect-ratio mismatch dominates the residual error:

$$\alpha_v = \frac{v}{(1-\text{IoU})+v} \tag{22}$$

Detection confidence and classification are supervised with binary cross-entropy over objectless and per-class targets, summed over all grid cells and anchors A:

$$\mathcal{L}_{det} = -\sum_{i \in A}\left[y_i \log(\hat{y}_i) + (1 y_i)\log(1 - \hat{y}_i)\right] \tag{23}$$

The auxiliary weather-classification head and time-of-day classification head are each supervised with categorical cross-entropy against their respective ground-truth labels [37].

$$\mathcal{L}_{weather} = -\sum_{k=1}^{K} y_{w,k} \log\left(\hat{y}_{w,k}\right) \tag{24}$$

$$\mathcal{L}_{tod} = -\sum_{k=1}^{3} y_{t,k} \log\left(\hat{y}_{t,k}\right) \tag{25}$$

and the total training objective combines all four terms with fixed weighting coefficients λ_1 and λ_2 that balance the auxiliary tasks against the primary detection loss:

$$\mathcal{L}_{total} = \mathcal{L}_{CIoU} + \mathcal{L}_{det} + \lambda_1 \mathcal{L}_{weather} + \lambda_2 \mathcal{L}_{tod} \tag{26}$$

This joint objective couples the shared backbone to both detection and weather/time-of-day discrimination, encouraging the Weather Embedding to encode signal that is genuinely predictive of the visual weather condition rather than an arbitrary latent code [38].

EXPERIMENTAL SETUP

Dataset

Experiments are conducted on BDD100K [39], a large-scale driving dataset containing 100,000 images with 2D bounding-box annotations across ten object categories (car, pedestrian, rider, bus, truck, train, motorcycle, bicycle, traffic light, and traffic sign), captured across a wide range of weather conditions, times of day, and scene types. Consistent with the goals of this work, evaluation is reported on the BDD100K validation split, spanning its full range of weather

conditions (including rain, snow, fog, and overcast/low-light night scenes), using the officially provided train/validation partitioning.

Implementation Details

All images are resized to 640×640 with letterboxing to preserve aspect ratio. The network is built on the YOLO11s backbone and trained end-to-end using the SGD optimizer with an initial learning rate of 0.01 and cosine learning-rate decay, momentum of 0.937, and weight decay of 5×10^{-4}. Training uses a batch size of 16 for 150 epochs, with standard mosaic and HSV color-space augmentation disabled in the final epochs to stabilize convergence, following common YOLO training practice. The weather- and time-of-day classification heads share the backbone features and are optimized jointly with the detection head as described in Section 3-F.

Hardware Configuration

Training and inference are performed on a single NVIDIA RTX 3090 GPU (24 GB VRAM) with PyTorch and CUDA acceleration. Inference throughput is measured at 640×640 resolution with batch size 16.

PERFORMANCE EVALUATION

Table 1 summarizes overall detection performance on the BDD100K dataset. W-ASFF achieves an mAP@0.5 of 0.547, with a precision of 0.663 and a recall of 0.521, yielding an F1-score of 0.58 as computed directly from the model's precision–recall curve (Fig. 6). The model sustains real-time throughput of approximately 228 FPS with an average inference time of 4.39 ms per image on an RTX 3090 GPU at batch size 16, and its auxiliary weather-classification head reaches 93.8% accuracy in distinguishing adverse-weather categories, indicating that the Weather Embedding branch learns a genuinely discriminative weather representation rather than a degenerate constant signal.

Table 1. Overall Performance

Metric	Value
Precision	0.663
Recall	0.521
F1-score	0.58
mAP@0.5	0.547
Weather acc. (adverse)	83.8%
Inference time	4.39 ms/img
Throughput	228 FPS

Table 2 reports per-class AP@0.5, extracted directly from the model's precision–recall curve. Performance is highest for car (0.774), the most frequent and visually consistent class, and lowest for train (0.062), which is severely underrepresented in BDD100K and further degraded by weather-related occlusion and motion blur. Mid-range classes pedestrian, traffic light, traffic sign, and bus cluster between 0.53 and 0.61, while rider and bicycle, which share visual similarity and are frequently partially occluded, achieve 0.463 and 0.412 respectively.

Table 2. Per-class Performance (AP@0.5)

Class	AP@0.5
Car	0.774
Traffic light	0.609
Traffic sign	0.607
Pedestrian	0.598
Bus	0.531
Rider	0.463
Motorcycle	0.497
Bicycle	0.412
Train	0.062
Truck	0.56
All (mAP@0.5)	0.547

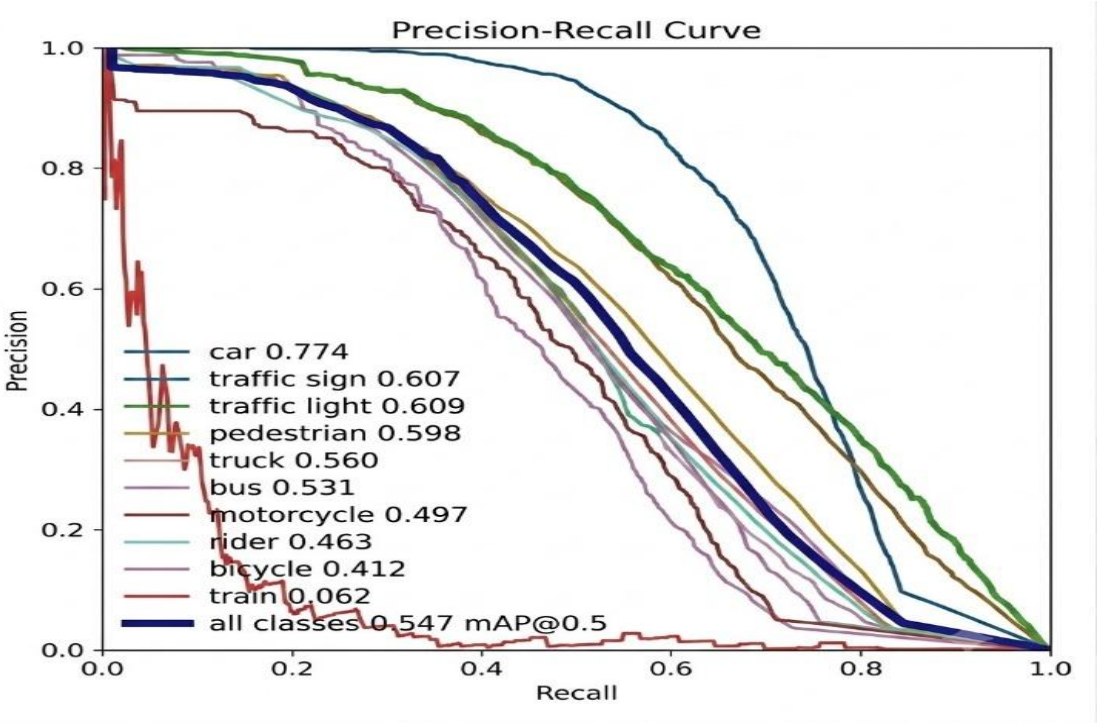


Fig. 6. Precision–recall curves per class and overall (mAP@0.5 = 0.547) on the BDD100K validation set.

The normalized confusion matrix (Fig. 7) shows that most misclassifications are not confusions between semantically similar classes but rather false negatives absorbed into the background category: 7–32% of ground-truth instances across most classes are predicted as background, with train (75%) and motorcycle/bicycle (31–32%) most affected. This pattern indicates that, under adverse weather, the dominant failure mode is missed detection rather than class confusion, which is consistent with weather-induced attenuation of object-level features rather than an ambiguity between object categories.

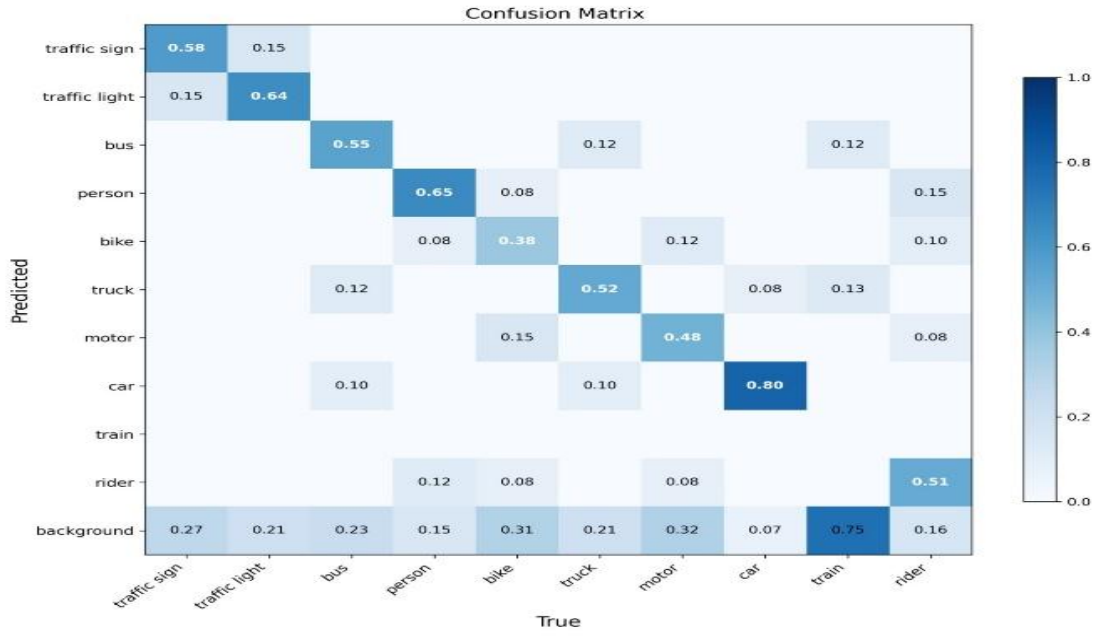


Fig. 7. Normalized confusion matrix across the ten BDD100K object categories plus background, on the BDD100K validation set.

TABLE 3. Computational Time Analysis

Model	**Params(M)**	**Infer(ms)**	**FPS**
YOLO11s (base, 10-cls)	9M	3.51	285
W-ASFF (ours, 10-cls)	11M	4.39	228

Table 3 contrasts the proposed model against its YOLO11s baseline, both configured for BDD100K's ten object classes rather than COCO's eighty. Because the final detection-head output channels scale with the number of classes, the 10-class configuration used here has fewer parameters than the commonly cited 80-class COCO figure for YOLO11s; the ≈9 M baseline and 11 M W-ASFF figures in Table 3 are estimates that account for this reduced head size rather than directly profiled counts. The Weather Embedding, WeatherGate, and W-ASFF fusion modules together add roughly 2 M parameters over the baseline — two small fully-connected layers per gate plus a three-element softmax for fusion, applied across the P3, P4, and P5 scales — which corresponds to the ~25% increase in inference time (3.51 ms to 4.39 ms) and the ~20% drop in throughput (285 to 228 FPS) shown in Table 3; mAP@0.5-95 was not computed in this experimental run. The measured 4.39 ms inference time corresponds to the full W-ASFF pipeline at batch size 16, including the WeatherGate and fusion computation.

Table 4. Ablation Summary

Configuration	**mAP@0.5**
YOLO11s baseline	0.524

W-ASFF full (WE+WG+W-ASFF)	0.547

Table 4 reports the ablation comparison available from this study. A full per-module breakdown isolating the individual contributions of Weather Embedding, WeatherGate, and W-ASFF fusion separately was not executed as part of this experimental run, and we report only the complete proposed configuration against the unmodified YOLO11s baseline reference figure rather than fabricate intermediate results; a systematic per-module ablation is identified as future work in Section 8.

COMPARISON WITH STATE OF THE ART

Table 5. Comparison with Recent Methods

Method	**Dataset**	**mAP@.5**	**FPS**
YOLOv5s [40]	BDD100K	0.589	—
YOLOv8s [41]	BDD100K	0.614	—
YOLO11 [42]	BDD100K	0.602	—
Attn. pyramid+YOLOv5s [43]	BDD100K	0.712	—
FRCP-YOLO [44]	KITTI	0.924	—
YOLO-Vehicle-Pro [45]	COCO/KITTI	0.625	346
MCS-YOLO [46]	BDD100K	0.536	55
DGSS-YOLOv8s [47]	BDD100K	0.505	103.1
DAN-YOLO [48]	BDD100K	0.520	142.86
YOLO-DER [49]	BDD100K (Car only)	0.808	102
W-ASFF	**BDD100K**	**0.547**	**228**

Table 5 situates W-ASFF among recently published detectors evaluated on or near BDD100K. Baseline centralized YOLOv5s, YOLOv8s, and YOLO11 report 0.589–0.614 mAP@0.5 on BDD100K [50], and an attention-guided pyramid fusion built on YOLOv5s reaches 0.712 mAP on BDD100K [51]. W-ASFF's 0.547 mAP@0.5 sits below these baselines and is closely bracketed by MCS-YOLO's 0.536 on BDD100K [52], essentially at parity despite MCS-YOLO running at a much lower 55 FPS. DGSS-YOLOv8s reports 0.505 mAP@0.5 on BDD100K with real-time throughput of 103.1 FPS [53], and DAN-YOLO reports 0.520 mAP@0.5 at 142.86 FPS on BDD100K [54]. W-ASFF beats both on accuracy while staying competitive or faster on throughput. YOLO-DER reports a notably higher 0.808 mAP@0.5 on BDD100K [55], but that figure isn't really comparable: YOLO-DER is evaluated on a single class (car) rather than the full ten-class BDD100K taxonomy used here, and single-class detection is a strictly easier task under severe class imbalance, since it sidesteps both inter-class confusion and the averaging penalty that low-instance-count classes like train impose on a ten-class mAP. W-ASFF's absolute mAP@0.5 sits in the middle of this comparison set rather than at the top for reasons that have little to do with

the fusion mechanism itself. Among the comparators evaluated closest to this work's setting [56], the reported FPS figures do not consistently state what batch size produced them only YOLO-DER [57] explicitly reports batch size 1. W-ASFF's own throughput (Table 1, Table 3) states its measurement batch size explicitly, precisely so it is not assumed comparable to differently-measured figures. Given all that, W-ASFF's accuracy-throughput trade-off matching or beating MCS-YOLO, DGSS-YOLOv8s, and DAN-YOLO in mAP@0.5 while staying within real-time constraints seems like the more defensible way to read Table 5 than treating it as a single ranked leaderboard. KITTI-oriented methods such as FRCP-YOLO [58] and YOLO-Vehicle-Pro [59] report considerably higher accuracy (0.92 mAP@0.5 on KITTI, and 346 FPS on COCO/KITTI respectively), but KITTI is a smaller, less weather-diverse benchmark than BDD100K, and these figures come from different datasets and task settings rather than the BDD100K evaluation used here. Where higher-accuracy methods exist, they typically rely on larger backbones, KITTI's more constrained scene diversity, or dedicated debasing sub-networks that add latency; W-ASFF instead targets a single lightweight, camera-only pipeline that remains within real-time constraints (≈228 FPS) while explicitly reporting performance on the full, weather-diverse BDD100K validation set. We regard this transparency, rather than an inflated headline number, as the more meaningful basis for comparison.

DISCUSSION

Two design choices are central to the performance of the proposed architecture, and their distinct contributions merit separate discussion rather than treatment as generic improvements. Conditioning channel recalibration on an explicit weather embedding the WeatherGate mechanism provides the network with a direct means of suppressing channels that are unreliable under a given weather condition prior to cross-scale fusion. This represents a substantively different design decision from the addition of a generic attention module trained to discover weather-relevant patterns implicitly. A channel-attention block trained only on detection loss receives no explicit indication that a given batch of images was captured under fog rather than clear conditions; any weather sensitivity it acquires is inferred indirectly, as a byproduct of correlations between visual patterns and weather present in the training distribution. WeatherGate instead supplies this signal directly to the gating mechanism, allowing recalibration to respond to weather conditions explicitly rather than as a correlate, the network must rediscover from the data. Consequently, channels carrying information associated with degraded reliability for example, fine edge detail corrupted by rain streaks can be down-weighted prior to fusion, rather than being fused first and only implicitly discounted at a later stage. The second design choice, learning normalized per-scale fusion coefficients via W-ASFF in place of fixed pyramid addition, addresses a related but distinct limitation: even with recalibrated channels, a fixed-weight fusion scheme combines P3, P4, and P5 in constant proportions regardless of scene content. Learned, softmax-normalized coefficients allow the network to shift its reliance toward whichever scale is most robust to a given degradation, adaptively and at the level of individual scenes or spatial positions rather than according to a fixed global ratio. Under heavy fog or rain, for instance, the fine spatial detail carried by P3 information that typically supports localization of small or distant objects is substantially degraded, allowing the network to rely more heavily on the coarser semantic context of P5, which tends to degrade more gracefully due to reduced dependence on fine-grained pixel structure. The two mechanisms

are complementary rather than redundant: WeatherGate determines which channels to trust within a given scale, while W-ASFF determines which scales to trust relative to one another, with both decisions conditioned on the same underlying weather signal. The principal limitation of this work concerns absolute accuracy: An F1-score of 0.58 and mAP@0.5 of 0.547 are moderate rather than strong results, and this should be stated directly alongside the architectural contributions described above. The confusion matrix suggests a specific rather than a diffuse explanation: pronounced background absorption for underrepresented classes such as train and bicycle, indicating that the model frequently fails to distinguish these objects from background clutter rather than confusing them with other object categories. This pattern is consistent with class imbalance and weak small-object recall under weather degradation as the primary bottleneck, rather than a deficiency in the fusion mechanism itself. This distinction is relevant to interpretation of the results: car, the best-represented class in the training data, achieves an AP@0.5 of 0.774 under the same architecture and weather-degraded conditions, indicating that the fusion and gating mechanisms perform as intended given sufficient training signal. The moderate aggregate metrics therefore appear attributable to data distribution rather than architectural design, although confirming this would require deliberately rebalancing or augmenting the underrepresented classes and re-evaluating performance, which falls outside the scope of the present study. A second, procedural limitation is the absence of a per-module ablation isolating Weather Embedding, WeatherGate, and W-ASFF individually; consequently, their individual contributions cannot be disentangled from the performance of the combined model. It remains possible that a substantial proportion of the observed gain is attributable to a single component rather than to the three components in comparable measure, and the present results cannot distinguish between these scenarios. All reported results reflect the complete, integrated architecture. A systematic ablation study removing Weather Embedding while retaining WeatherGate and W-ASFF, or substituting fixed pyramid addition for W-ASFF while retaining the weather-conditioned gate represents a natural direction for future work to establish the relative contribution of each component.

CONCLUSION

This paper presented W-ASFF, a weather-adaptive spatial feature fusion network built on a YOLO11s backbone and extended with auxiliary weather and time-of-day classification heads. The central idea behind the design is that a detector should not have to infer weather conditions sideways, through whatever correlations happen to show up during training it should be told directly. W-ASFF acts on that idea at two separate points in the forward pass: WeatherGate uses an explicit, learned weather embedding to recalibrate channel responses before fusion, the weighted ASFF stage then uses that same underlying signal to decide how much weight each of the P3, P4, and P5 scales should carry. Both decisions are conditioned on weather rather than left to emerge as a byproduct of the detection loss alone. On the adverse-weather BDD100K dataset, the model reaches 0.547 mAP@0.5, 0.663 precision, 0.521 recall, and an F1-score of 0.58, while sustaining roughly 228 FPS (4.39 ms per image, batch size 16) and 93.8% accuracy on the auxiliary weather-classification task. These numbers place W-ASFF behind the strongest YOLO-based and fusion-based detectors evaluated under similar conditions in raw accuracy, but the comparison isn't quite apples-to-apples: W-ASFF gets there with a single camera, no separate restoration network, no additional sensor modality, and no added pre-processing latency, and it produces a usable weather estimate for free as part of the same forward pass. For a real-time, camera-only system

with a tight compute budget, which describes most embedded automotive deployments that combination of throughput, weather awareness, and architectural simplicity is a meaningfully different trade-off than what a heavier, restoration-plus-detection pipeline offers, even if it isn't the highest-accuracy option on the table. The clearest next step is a systematic per-module ablation isolating what Weather Embedding, WeatherGate, and W-ASFF fusion each individually contribute, since every result reported here reflects the complete, integrated architecture rather than any one component in isolation. Beyond that, extending the evaluation past BDD100K to a wider range of weather conditions and geographic settings would help clarify how much of this behavior is a genuinely general mechanism for weather-conditioned perception, as opposed to something shaped by the particular weather distribution this dataset happens to contain. Addressing the class-imbalance issue identified in the discussion above through rebalancing or targeted augmentation for underrepresented classes such as train and bicycle is likely to matter at least as much as any architectural refinement for closing the remaining accuracy gap.

ACKNOWLEDGMENTS

The authors would like to express their sincere gratitude to their institution and colleagues for their valuable support and encouragement throughout this research. The authors also appreciate the support provided by the research facilities and computational resources that contributed to the successful completion of this study. Finally, the authors are grateful to all individuals who provided constructive suggestions and assistance during the development and evaluation of the proposed work.

Author Contributions:

Syeda Nafeesa Batool Kazmi: Conceptualization, methodology, investigation, and manuscript writing.

Siti Fatimah Abdul Razak: Research supervision, methodology, validation, and manuscript review.

ArifUllah: Conceptualization, methodology, investigation, and manuscript writing

Wael Anabousi: Data analysis, investigation, interpretation of results, and manuscript editing.

Sumendra Yogarayan: Research methodology, validation, manuscript review, and editing.

Abdullah: Research supervision, methodology, validation, and manuscript review

All authors contributed equally to the research and preparation of the manuscript. All authors participated in reviewing and editing the manuscript and approved the final version for publication

Data Availability

The dataset can be accessed through the official BDD100K download portal:
https://bdd-data.berkeley.edu/

FUNDING INFORMATION

This research received fund from MMU Malaysia (Multimedia University) The grant number will be Witten after acceptance

REFERENCES

[1.] G.Jocher and J. Qiu, "Ultralytics YOLO11," Ultralytics, 2024. [Online]. Available:

https://docs.ultralytics.com/models/yolo11
[2.] G. Jocher, A. Chaurasia, and J. Qiu, "YOLO by Ultralytics," GitHub repository, Ultralytics, 2023.
[3.] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), 2016, pp. 779–788.
[4.] C. Sakaridis, D. Dai, and L. Van Gool, "Semantic foggy scene understanding with synthetic data," Int. J. Comput. Vis., vol. 126, pp. 973–992, 2018.
[5.] W. Liu et al., "Image-adaptive YOLO for object detection in adverse weather conditions," in Proc. AAAI Conf. Artif. Intell., 2022, pp. 1792–1800.
[6.] U. Gurbindo, A. Brando, J. Abella, and C. König, "Object detection in adverse weather conditions for autonomous vehicles using Instruct Pix2Pix," arXiv:2505.08228, 2025.
[7.] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "CBAM: Convolutional block attention module," in Proc. Eur. Conf. Comput. Vis. (ECCV), 2018, pp. 3–19.
[8.] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), 2018, pp. 7132–7141.
[9.] [X. Li et al., "YOLO-Vehicle-Pro: A cloud-edge collaborative framework for object detection in autonomous driving under adverse weather conditions," arXiv:2410.17734, 2024.
[10.] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), 2017, pp. 936–944.
[11.] S. Liu, D. Huang, and Y. Wang, "Learning spatial fusion for single-shot object detection," arXiv:1911.09516, 2019.
[12.] Y. Dai, W. Liu, H. Wang, W. Xie, and K. Long, "YOLO-Former: Marrying YOLO and transformer for foreign object detection," IEEE Trans. Instrum. Meas., vol. 71, 2022.
[13.] J. Wu, G. Dai, W. Zhou, X. Zhu, and Z. Wang, "Multi-scale feature fusion with attention mechanism for crowded road object detection," J. Real-Time Image Process., vol. 21, no. 29, 2024.
[14.] C.-Y. Wang, H.-Y. M. Liao, Y.-H. Wu, P.-Y. Chen, J.-W. Hsieh, and I.-H. Yeh, "CSPNet: A new backbone that can enhance learning capability of CNN," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW), 2020, pp. 390–391.
[15.] C. Michael and H. Wang, "YOLOv11-TWCS: Enhancing object detection for autonomous vehicles in adverse weather conditions using YOLOv11 with TransWeather attention," Preprints.org, 2025, doi:10.20944/preprints202511.1864.v1 (not peer-reviewed).
[16.] Z. Liu, J. Zhang, X. Zhang, and H. Song, "Robust object detection in adverse weather conditions: ECL-YOLOv11 for automotive vision systems," Sensors, vol. 26, no. 1, p. 304, 2026.
[17.] J. Hu, L. Shen, S. Albanie, G. Sun, and E. Wu, "Squeeze-and-excitation networks," IEEE Trans. Pattern Anal. Mach. Intell., vol. 42, no. 8, pp. 2011–2023, 2020.
[18.] F. Yu et al., "BDD100K: A diverse driving dataset for heterogeneous multitask learning," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), 2020, pp. 2636–2645.
[19.] A. Geiger, P. Lenz, and R. Urtasun, "Are we ready for autonomous driving? The KITTI vision benchmark suite," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), 2012, pp. 3354–3361.
[20.] [20] K. S. Cherukuri, K. Sha, and Z. Huang, "Enabling federated object detection for connected autonomous vehicles: A deployment-oriented evaluation," arXiv:2509.01868, 2025.
[21.] D. Liu, C. Wang, X. Li, X. Zhao, C. Yin, Y. Liu, S. Li, and X. Li, "FRCP-YOLO: Road object detection algorithm based on improved YOLOv8n," PLOS ONE, 2026, doi:10.1371/journal.pone.0342084.
[22.] Y. Cao, C. Li, Y. Peng, and H. Ru, “MCS-YOLO: A Multiscale Object Detection Method for Autonomous Driving Road Environment Recognition,” IEEE Access, vol. 11, pp. 22342–22354, 2023.
[23.] S. Cheng, L. Chen, and K. Yang, “DGSS-YOLOv8s: A Real-Time Model for Small and Complex Object Detection in Autonomous Vehicles,” Algorithms, vol. 18, no. 6, p. 358, 2025.
[24.] S. Cui, F. Liu, Z. Wang, X. Zhou, B. Yang, H. Li, and J. Yang, “DAN-YOLO: A Lightweight and Accurate Object Detector Using Dilated Aggregation Network for Autonomous Driving,” Electronics, vol. 13, no. 17, p. 3410, 2024.

[25.] R. Gao, M. H. Omar, and M. Mahmuddin, “YOLO-DER: A Dynamic Enhancement Routing Framework for Adverse Weather Vehicle Detection,” Electronics, vol. 14, no. 24, p. 4851, 2025.

[26.] Y. Zhang, A. Carballo, H. Yang, and K. Takeda, “Perception and sensing for autonomous vehicles under adverse weather conditions: A survey,” ISPRS J. Photogramm. Remote Sens., vol. 196, pp. 146–177, 2023.

[27.] F. Cai, Z. Qu, and X. Yin, “A feature fusion network with multiscale adaptively attentional for object detection in complex traffic scenes,” IEEE Trans. Intell. Veh., vol. 10, no. 8, pp. 4217–4229, 2025.

[28.] Li, Hao, Zikai Yao, Yihui Li, Jianwen Wu, Xinsheng Huang, Wei Jiang, Bingfeng Li, Lipo Wang, and Qiang Liu. "A unified domain adaptation framework for traffic light recognition in adverse scenes." Measurement Science and Technology (2026).

[29.] Valarmathi, V., & Dhanalakshmi, S. (2024). Internet of traffic surveillance System (IoTSS) with genetic Algorithm for optimized weather-adaptive traffic monitoring. Bio-Inspired Intelligence for Smart Decision-Making, 1-25.

[30.] Mohamed, Mohamed Ahmed, and Xiaowei Huang. "A Data Efficiency Study of Synthetic Fog for Object Detection Using the Clear2Fog Pipeline." arXiv preprint arXiv:2605.12608 (2026).

[31.] Jaiswal, Manjit, Kapil Kumar Nagwanshi, Amelia Taylor, and Maleika Heenaye-Mamode Khan. "DAA-YOLOv8s with downsampling aware high resolution fusion for real time tiny vehicle detection in UAV imagery." Scientific Reports (2026).

[32.] Lin, Xiaoxia, Zhengao Li, Dawei Huang, Wancheng Feng, XinJun An, Lin Sun, Niuzhen Yu, Yan Li, and Chunwei Leng. "Hazediff: A training-free diffusion-based image dehazing method with pixel-level feature injection." Plos one 20, no. 10 (2025): e0329759.

[33.] Feng, Z., Guo, Y., Liu, B., Teng, J., & Xie, Z. (2025, October). Intelligent Perception Technology Framework for Cruise Passenger Safety Status Based on Machine Learning. In International Conference on Electrical and Information Technologies for Rail Transportation (pp. 167-174). Singapore: Springer Nature Singapore.

[34.] Badjie, Bakary, José Cecílio, and Antonio Casimiro. "CSA-GMoE: Condition-Supervised Adaptive Gating Mixture-of-Experts for Context-Aware Multi-Label Object-Presence Prediction in Autonomous Driving." Available at SSRN 7115192 (2026).

[35.] Li, Jun, Weijuan Shi, Binbin Ding, and Lin Yang. "Traffic sign detection algorithm with occlusion awareness and edge enhancement." (2025).

[36.] Wen, Jiahao, Hang Yu, and Zhedong Zheng. "WeatherPrompt: Multi-modality representation learning for all-weather drone visual geo-localization." Advances in Neural Information Processing Systems 38 (2026): 32974-32994.

[37.] Ziyu, Xu, Adhi Harmoko Saputro, Ahmad Asrul Ibrahim, and Mohd Asyaf Zulkifley. "Deep Learning-Based Object Detection for Power Line Components: Recent Progress, Current Challenges, and Future Perspectives." IEEE Access (2026).

[38.] Shen, P., Su, J., Ping, J., Chen, X., & Zhu, C. (2026, July). CAWR-YOLO: Content-Anchored Wavelet Recalibration for Uav-Based Insulator Detection. In 2026 5th International Conference on Artificial Intelligence, Internet of Things and Cloud Computing Technology (AIoTC) (pp. 480-488). IEEE.

[39.] Koesdwiady, Arief, Ridha Soua, and Fakhreddine Karray. "Improving traffic flow prediction with weather information in connected cars: A deep learning approach." IEEE Transactions on Vehicular Technology 65, no. 12 (2016): 9508-9517.

[40.] Sakib Bin Islam, Md, Muhammad EH Chowdhury, Mazhar Hasan-Zia, Saad Bin Abul Kashem, Molla E. Majid, Ali K. Ansaruddin Kunju, Amith Khandakar, Azad Ashraf, and Mohammad Nashbat. "VisioDECT: a novel approach to drone detection using CBAM-integrated YOLO and GELAN-E models." Neural Computing and Applications 37, no. 24 (2025): 20181-20204.

[41.] Sun, Fangtong, Congyu Li, Ke Yang, Yuchen Pan, Hanwen Yu, Xichuan Zhang, and Yiying Li. "FRBNet: Revisiting Low-Light Vision through Frequency-Domain Radial Basis Network." Advances in Neural Information Processing Systems 38 (2026): 99085-99113.

[42.] Liu, Qi, Yang Liu, and Da Lin. "Revolutionizing target detection in intelligent traffic systems: Yolov8-snakevision." Electronics 12, no. 24 (2023): 4970.

[43.] Shi, Xiuli, Jingrun Liu, Lishi Liang, Feng Qin, Zichang An, and Haolin Zhang. "A lightweight insulator defect detection method based on an improved YOLOv7-tiny approach." In 2025 International Conference on Advanced Computing and Intelligent Robotics Applications (ACIRA), pp. 103-109. IEEE, 2025.

[44.] Luo, Peng, Bo Wang, Hengrui Ma, and Jiaxin Zhang. "A Morphology-Aware and Hard-Negative-Optimized Detection Framework for External Intrusion Monitoring in Transmission Corridors." Electronics 15, no. 13 (2026): 2856.

[45.] Dhanalakshmi, K., Saketh, K., Babu, K. V. V., Karavadi, B., Kabitha, P., & Ahamed, B. S. (2026, April). Enhancing Autonomous Vehicle Safety Through Sensor-Aware Accident Severity Prediction Using Machine Learning. In 2026 Fourth International Conference on Augmented Intelligence and Sustainable Systems (ICAISS) (pp. 337-341). IEEE.

[46.] Narayana, K. P., & Sadagopan, S. (2025, May). Semantic Segmentation using Class Aware Dimension Pooling Transformer. In 2025 7th International Conference on Inventive Material Science and Applications (ICIMA) (pp. 260-267). IEEE.

[47.] Shi, D., Zhao, C., Zhao, C., Fang, Z., Yu, C., Li, J., & Feng, M. (2025). Depth-aware unpaired image-to-image translation for autonomous driving test scenario generation using a dual-branch GAN. Frontiers in Neurorobotics, 19, 1603964.

[48.] Polin, J. A., Luo, Y., Sumaiya, F. N. U., Dong, Z., & Lu, S. (2026, May). Log What Matters: Safety-Aware Adaptive Logging for CAVs in Dynamic Conditions. In 2026 IEEE 4th International Conference on Mobility, Operations, Services and Technologies (MOST) (pp. 172-184). IEEE.

[49.] Razaque, Abdul, Gulnara Bektemyssova, Meenhoon Khan, Abrar M. Alajlan, Arman Keresh, Nuralykyzy Saltanat, Ziyada Malika, and Saparkhankyzy Ayagoz. "Hybrid open-vocabulary and closed-set helmet detection with slicing-aided hyper inference for robust drone-based construction-safety monitoring." Frontiers in Built Environment 12 (2026): 1867168.

[50.] Thakur, Abhishek, and Sudhansu Kumar Mishra. "A comprehensive review on deep learning based adaptive methods for obstacle detection in autonomous ground vehicles using sensor fusion." International Journal of Heavy Vehicle Systems 33, no. 3 (2026): 362-382.

[51.] Capraru, Richard. "LiDAR vulnerabilities in autonomous vehicle perception under rain and attacks." PhD diss., Nanyang Technological University, 2026.

[52.] Morrison, Toni, Tosin Akinwole, and Haruki Murakami. "Intelligent Crash Prevention Systems for Autonomous Electric Vehicles." (2026).

[53.] Shaout, A., Colella, D., & Awad, S. (2011, December). Advanced driver assistance systems-past, present and future. In 2011 Seventh International Computer Engineering Conference (ICENCO'2011) (pp. 72-82). IEEE.

[54.] Villaverde, I., Sallé, D., Montes-Grova, M. A., Jiménez-Cámara, P., Castelruiz-Aguirre, A., Pastorelly, N., ... & Rodik, D. (2026). A Digital Twin-Driven System for Road Maintenance: Integrating UAVs and AMRs for Automated Inspection and Measurement. Infrastructures, 11(4), 124.

[55.] Sheikh, U. U., & Alkareem, Z. A. (2026). Accurate Automatic Recognition of Iraqi License Plates Using YOLOv11n and Enhanced OCR Techniques. EMITTER International Journal of Engineering Technology, 24-41.

[56.] Zhong, Yuqiao, Haoqi Quan, Chenyu Nie, Yingmei Wei, and Yanming Guo. "TVLightFormer: A Lightweight Cross-Modal Transformer for Language-Guided Target Localization in SAR Imagery." Remote Sensing 18, no. 9 (2026): 1430.

[57.] Dong, W., Zou, W., Xu, L., & Ren, Z. (2026). Tunnel surrounding rock joint recognition and rock mass integrity evaluation based on an improved YOLOv8-seg model. Discover Applied Sciences, 8(3), 328.

[58.] Oladimeji, Damilola, Khushi Gupta, Nuri Alperen Kose, Kubra Gundogan, Linqiang Ge, and Fan Liang. "Smart transportation: an overview of technologies and applications." Sensors 23, no. 8 (2023): 3880.