

Paper Type: Original Article

Cross-Domain Computer Vision Model for Adverse Weather Image Recognition

Yahya Albugg¹ , Maan Y Anad Alsaleem²  and Omar Shakir Hasan^{2,*} 

¹ Northern Technical University, Mosul, Iraq; dr.yahya.albugg@ntu.edu.iq.

² Directorate of Education in Nineveh, Mosul, Iraq;
Emails: maanyounis1983@gmail.com; omarshakir06@gmail.com.

Received: 11 Dec 2025

Revised: 19 Mar 2026

Accepted: 24 Jun 2026

Published: 27 Jun 2026

Abstract

Real-world computer vision systems are affected by various types of weather conditions, such as fog, rain, snow, haze, etc.; however, the models trained on clean images can't generalize well to real-world images due to the significant change of visual appearance under different weather conditions. Most of the existing methods either treat weather removal and visual recognition as two separate tasks or focus on a single type of degradation at a time, which is less applicable to practical scenarios where a variety of adverse weather may exist; therefore, in this paper, we propose a unified CDW-Net (Cross-Domain Weather-Aware Vision Network) for image recognition under various challenging weather conditions. The proposed CDW-Net employs a hybrid architecture consisting of a degradation-aware transformer encoder and a convolutional feature refinement decoder, which enables both understanding and classification of weather-degraded images in one framework. Specifically, the network contains two main branches, including a global self-attention branch built on a modified Swin Transformer to capture the long-range degradation patterns caused by weather and a local convolutional branch to capture the fine-grained structural details that remain visible even if parts of the scene are occluded or blurred. The two streams of features are then combined by a Cross-Domain Attention Fusion (CDAF) module, which also applies the Maximum Mean Discrepancy (MMD) regularization to better align the feature distributions of clean (source) and degraded (target) domains for mitigating the domain shift. Consequently, we evaluate the proposed CDW-Net on three benchmark datasets, i.e., Multi-class Weather Dataset (MWD), Weather Phenomenon Database (WEAPD), and Detection in Adverse Weather Nature (DAWN) dataset, where CDW-Net achieves top-1 accuracies of 97.84%, 91.23%, and 89.67%, respectively, outperforming all the state-of-the-art baselines. Ablation studies indicate that each of the architectural components contributes meaningfully to the final performance; moreover, our proposed framework is also scalable and practical for deployment in real-world all-weather visual recognition applications, such as autonomous driving, surveillance, and intelligent transportation systems.

Keywords: CDW-Net; Cross-Domain Attention Fusion; Maximum Mean Discrepancy; Weather Conditions.

1 | Introduction

The generalization of computer vision models across different atmospheric conditions is an essential factor to determine their reliability in the real-world environment, although the current deep learning models have approached near-human accuracy on benchmark clean-image datasets. The accuracy of these models drops



Corresponding Author: omarshakir06@gmail.com



<https://doi.org/10.61356/j.nois.2026.10661>



Licensee **Neutrosophic Optimization and Intelligent systems**. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0>).

significantly when the images are captured in adverse weather conditions, such as fog, rain, snow, haze, and others, which presents a major safety risk to many real-world applications, including autonomous driving, drone-based surveillance, intelligent transportation systems, and outdoor environmental monitoring, where models are expected to provide stable and accurate recognition of objects in different atmospheric conditions [1, 2].

Weather conditions degrade image quality and fog and haze reduce visibility and contrast, rain creates streaks and reflections on illuminated surfaces, snow causes overexposure and obscures details [3]. All these phenomena distort appearance in various ways, which are not invariant across scenes, therefore this causes appearance variations between clear images (used for training) and degraded ones (seen at test time). This phenomenon has been recognized and it's why models trained on clean images perform significantly worse on images with adverse weather [4].

There are two approaches to address this problem, and the first one is image restoration, which aims at restoring the image first and then feeding the restored image to a recognition model [5, 6]. Methods such as DehazeFormer [5] and TransWeather [7] can produce high-quality restored images, because the separation of image restoration and recognition usually leads to error accumulation over time and increases the cost. The second one is named as domain adaptation, and this approach aims at aligning the feature distributions of the source and target domains, which allows the model to generalize well when the input image is blurry [8]. Domain adaptation achieves promising results under certain circumstances, however, most existing approaches are tailored for a single type of adverse weather conditions, which restricts their applicability in handling multiple and mixed degradations in practice [9].

However, a more critical and less investigated problem is that real-world adverse weather is rarely simple or isolated, because most real-world autonomous vehicles and outdoor cameras have to operate in mixed weather, such as rain mixed with haze, or snow mixed with fog, where different degradation patterns interact and amplify, which cannot be well handled by a single-weather model [4]. Although there exist multi-weather restoration networks, e.g., MWFormer, they mainly focus on improving the image quality rather than addressing the semantic classification directly, meanwhile, most existing weather image classification methods do not explicitly model or correct the domain shift, but employ a deeper network or data augmentation for limited improvement [10, 11], which generalizes poorly to unseen or more complex weather-induced degradations.

In this paper, we present the Cross-Domain Weather-Aware Vision Network (CDW-Net), an end-to-end deep learning model that can recognize images captured in bad weather without requiring an additional restoration step prior to recognition. The proposed CDW-Net model directly takes the weather-degraded images as inputs and is composed of a dual-branch encoder, which employs a Swin Transformer to model the global weather-induced degradation patterns, and a CNN to learn the local structural details that are still perceivable from the weather-degraded images, and the two branches are then fused by a Cross-Domain Attention Fusion (CDAF) module that adaptively aligns features from the clean and degraded domains via MMD-based domain regularization. By modeling global degradation cues and local semantic structure jointly, CDW-Net is capable of generalizing to multiple weather conditions simultaneously, by exploiting semantic features that are still reliable even under severe visual corruption.

The principal contributions of this work are:

- CDW-Net architecture: A novel dual-branch encoder-decoder model combining global transformer-based attention with local CNN-based feature extraction for multi-weather image recognition.
- Cross-Domain Attention Fusion (CDAF): A learned fusion module incorporating MMD-based domain alignment to reduce distributional shift between clean-weather source domains and adverse-weather target domains.

- Multi-dataset evaluation: Comprehensive evaluation on MWD, WEAPD, and DAWN benchmarks demonstrating consistent state-of-the-art performance across diverse degradation types and scene categories.
- Ablation analysis: Detailed component-level ablation studies validating the contribution of each architectural block.

The remainder of this paper is organized as follows: Section 2 reviews the related work in three categories, weather image restoration, weather image classification, and domain adaptation for visual recognition, and Section 3 introduces the proposed CDW-Net architecture and its mathematical formulation. Section 4 reports the experimental setup, including datasets, implementation details, and evaluation metrics and results. Section 5 discusses the major observations, analyzes typical failure cases, and lists the current limitations of our approach. Section 6 concludes this paper and indicates potential directions for future studies.

2 | Related Works

2.1 | Adverse Weather Image Restoration

Research in image restoration in bad weather has been ongoing for many years and has transitioned from hand-crafted and physics-based models to the modern deep learning methods. Traditional approaches, such as the dark channel prior, assume a set of simplistic statistics about haze-free images, which is used to estimate the scene transmission map, however, when the assumptions are not valid, such methods perform poorly. In contrast, the encoder-decoder convolutional networks significantly enhance the restoration quality, since they can learn complex degradation patterns directly from the data, rather than depending on fixed priors.

The emergence of vision transformers also opens up a new frontier in image restoration tasks by modeling the long-range spatial dependencies beyond the receptive field of traditional CNN architectures. For example, Song et al. [5] proposed the dehazing method DehazeFormer based on Swin Transformer backbone, with custom normalization layers and improved spatial information aggregation, and the proposed method achieved a PSNR of over 40 dB on the SOTS indoor dataset and consistently surpassed all the existing CNN-based approaches. According to the survey from Ali et al. [6], the transformer-based methods are more and more becoming the preferred solution for a variety of image restoration tasks, such as dehazing, denoising, and weather removal, because they can model long-range dependencies more effectively than traditional CNN architectures.

The multi-weather restoration has also been explored recently. Valanarasu et al. [7] proposed the TransWeather, which is a transformer-based architecture with a single encoder and learnable weather-type embeddings to unify the removal of rain, fog, and snow into one framework. Based on the TransWeather [7], Zhu et al. [4] proposed the MWFormer, which utilizes hyper-networks and feature-wise linear modulation to adaptively adjust restoration parameters under multiple concurrent weather conditions, and recently, Özdenizci et al. [12] proposed the denoising diffusion probabilistic model for the patch-based image restoration, which achieves impressive generalization to the real-world degraded images in zero-shot settings. Although these methods have greatly improved the visual quality, they only focus on restoration and do not take into account the downstream image classification task, and therefore, the joint design of integrated restoration-and-recognition frameworks is still an open research problem.

2.2 | Weather Image Classification

Weather classification from images has also recently gained popularity as a computer vision problem in its own right, with obvious connections to applications like autonomous driving and environmental

monitoring. Albelwi [13] proposed a weather recognition method using DenseNet-121 with transfer learning and geometric data augmentation, achieving an accuracy of 95.9% on the RFS dataset, thus establishing a strong CNN-based baseline. The study also emphasized the importance of both network depth and the choice of augmentation strategy for achieving high performance on weather classification tasks.

Li et al. [11] proposed a hybrid weather classification framework with pre-trained Vision Transformer (ViT) as the backbone, and they also employed two augmented attention modules, convolutional self-attention and atrous self-attention. The authors achieved F1 scores of 97.47% and 87.69% on MWD and WEAPD datasets, respectively. Yıldırım et al. [14] proposed a weather classification method by concatenating the features obtained from MobileNetV2, DenseNet201, and EfficientNetB0, and feeding the concatenated features to an SVM classifier, and they used the mRMR feature selection method. Their results showed that feature fusion from multiple architectures outperforms the single pre-trained models. Chen et al. [10] designed a multi-label weather recognition model, MASK-CNN-Transformer (MASK-CT), which integrates convolutional feature extraction, transformer encoder, and mask-based regularization, and achieved state-of-the-art results on real-world weather datasets. Although the above classifiers have shown competitive results on their own datasets, they do not have an explicit domain alignment mechanism, which is why their generalization performance is often compromised when the training and test distributions are significantly different, and this is exactly the gap filled by our work.

Afxentiou et al. [15] provided a detailed benchmark study on CNN-based classification of adverse weather images for autonomous driving, and their experimental results showed that all architectures tested consistently demonstrated performance degradation in the challenging multi-class weather conditions scenario, which underscored the need for architectures that are robust to domain shift caused by adverse weather.

2.3 | Domain Adaptation for Visual Recognition

Domain adaptation methods aim to align the feature distributions of the labeled source domain with the unlabeled or label-scarce target domain. Hu et al. [16] proposed a domain-adaptive Faster R-CNN called DAGL-Faster, which incorporates both global-local image-level and instance-level domain classifiers for vehicle detection in foggy and rainy conditions, and their DAGL-Faster model achieves 36.7% mAP on the Foggy Cityscapes dataset. Yuan et al. [9] tested five standard unsupervised domain adaptation methods for haze image recognition and showed that aligning the distributions of source (clean) and target (hazy) domains leads to substantially higher classification accuracy than naive transfer. Goel et al. [8] proposed a guided transfer learning framework that minimizes JS-Divergence for domain-adaptive image classification and object detection. The model obtains 51.1% mAP on Foggy Cityscapes.

Recent image classification works on deep learning, including systematic reviews of deep learning-based approaches [17] and multi-ViT ensemble methods for fine-grained recognition [18], underscore that combining transformer-based attention with domain-aware training is a promising direction for improving generalization in distribution-shifted settings. The gap in the existing literature is the absence of a unified, end-to-end architecture that combines multi-weather recognition, domain-adaptive feature alignment, and hybrid global-local feature extraction within a single deployable framework — precisely the contribution of CDW-Net.

3 | Proposed Methodology

3.1 | Overview

CDW-Net is a dual-branch encoder-decoder architecture designed for end-to-end adverse weather image recognition. Given an input image $\mathbf{x} \in \mathbb{R}^{H \times W \times 3}$ captured under an arbitrary weather condition, CDW-Net

predicts the weather category $\hat{y} \in \{1, \dots, C\}$ where C is the number of weather classes. The architecture consists of four main components: (1) a Transformer Branch for global degradation-pattern modeling, (2) a Convolutional Branch for local structural feature extraction, (3) a Cross-Domain Attention Fusion (CDAF) module for adaptive feature alignment and fusion, and (4) a Classification Head producing the final prediction.

Figure 1 show Overview of CDW-Net. The input image is processed in parallel by a Swin Transformer branch (global features) and a ResNet-50 convolutional branch (local features). Features are fused via the Cross-Domain Attention Fusion (CDAF) module with MMD-based domain alignment. The classification head produces the final weather category label.

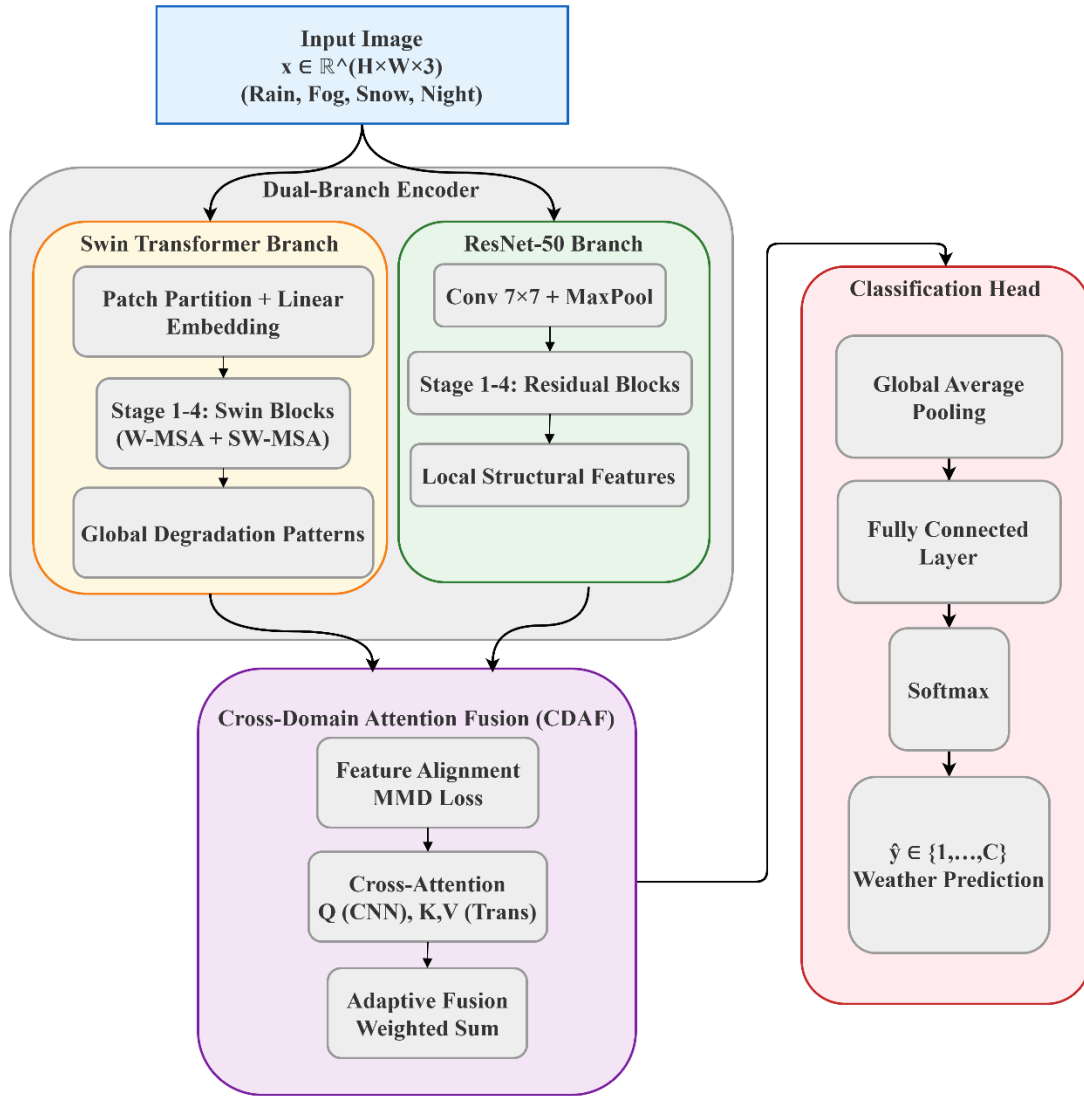


Figure 1. CDW-Net architecture overview.

3.2 | Transformer Branch

The transformer branch employs a Swin Transformer [5] backbone as the global encoder. Swin Transformer uses hierarchical shifted window-based self-attention to capture long-range spatial dependencies efficiently at multiple scales. Let the input feature map at stage l be $F_l^T \in \mathbb{R}^{\frac{H}{2^{l+1}} \times \frac{W}{2^{l+1}} \times C_l}$. The window-based multi-head self-attention (W-MSA) at stage l computes:

$$\text{Attention}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{softmax}\left(\frac{\mathbf{Q}\mathbf{K}^\top}{\sqrt{d_k}} + \mathbf{B}\right) \mathbf{V}$$

where $\mathbf{Q}, \mathbf{K}, \mathbf{V} \in \mathbb{R}^{n \times d_k}$ are the query, key, and value matrices computed from patch tokens within each window of size $M \times M$, d_k is the key dimension, and $\mathbf{B}$ is a learnable relative position bias matrix. The Shifted Window Multi-head Self-Attention (SW-MSA) alternates with W-MSA to enable cross-window information exchange, capturing weather-induced global degradation patterns that span large image regions.

The transformer branch outputs a feature tensor $\mathbf{F}^T \in \mathbb{R}^{\frac{H}{32} \times \frac{W}{32} \times C_T}$ from stage 4, which is subsequently global average pooled to yield a global descriptor $\mathbf{g}^T \in \mathbb{R}^{C_T}$.

3.3 | Convolutional Branch

The convolutional branch uses a ResNet-50 backbone for local feature extraction. ResNet-50 employs residual connections to facilitate gradient flow in deep networks and captures hierarchical local features through learned convolutional kernels. Let $\mathbf{F}^C \in \mathbb{R}^{\frac{H}{32} \times \frac{W}{32} \times C_C}$ denote the output of the final ResNet-50 convolutional block. The residual block at layer l computes:

$$\mathbf{F}_{l+1}^C = \text{ReLU}(\mathbf{F}_l^C + \mathcal{F}(\mathbf{F}_l^C; \mathbf{W}_l))$$

where $\mathcal{F}(\cdot; \mathbf{W}_l)$ denotes the stacked convolution-batch normalization-ReLU transformation, and this block performs very well in preserving edges and textures in regions of the image that are less affected by partial weather occlusion. This block is suitable for the transformer's global damage model because it performs well in preserving edges and textures in regions of the image that are less affected by partial weather occlusion, and thus Global average pooling is applied after the final residual block to obtain the local descriptor $\mathbf{g}^C \in \mathbb{R}^{C_C}$.

3.4 | Cross-Domain Attention Fusion (CDAF)

The CDAF module integrates features from both branches while reducing domain discrepancy between source and target distributions. Let $\mathbf{g}^T$ and $\mathbf{g}^C$ be the global descriptors from the transformer and convolutional branches, respectively. These are first concatenated and projected:

$$\mathbf{g}^{\text{cat}} = [\mathbf{g}^T \parallel \mathbf{g}^C] \in \mathbb{R}^{C_T + C_C}$$

$$\mathbf{g}^{\text{proj}} = \text{LayerNorm}(\text{Linear}(\mathbf{g}^{\text{cat}})) \in \mathbb{R}^d$$

A cross-attention mechanism computes attention weights between the two branch descriptors:

$$\alpha = \text{softmax}\left(\frac{(\mathbf{W}_Q \mathbf{g}^T)(\mathbf{W}_K \mathbf{g}^C)^\top}{\sqrt{d}}\right)$$

$$\mathbf{g}^{\text{fused}} = \alpha \cdot (\mathbf{W}_V \mathbf{g}^C) + (1 - \alpha) \cdot (\mathbf{W}_V \mathbf{g}^T)$$

where $\mathbf{W}_Q, \mathbf{W}_K, \mathbf{W}_V \in \mathbb{R}^{d \times d}$ are learned projection matrices. The adaptive weighting α allows the network to dynamically emphasize global transformer features for highly degraded images while retaining local convolutional features for partially occluded inputs.

Domain Alignment via MMD: During training with labeled source domain samples $\{(\mathbf{x}_s, \mathbf{y}_s)\}$ and unlabeled target domain samples $\{\mathbf{x}_t\}$, the CDAF module minimizes the Maximum Mean Discrepancy (MMD) between source and target feature distributions:

$$\mathcal{L}_{\text{MMD}} = \left\| \frac{1}{n_s} \sum_{i=1}^{n_s} \phi(g_i^{\text{fused}}) - \frac{1}{n_t} \sum_{j=1}^{n_t} \phi(g_j^{\text{fused}}) \right\|_{\mathcal{H}}^2$$

where $\phi(\cdot)$ is the kernel embedding in a reproducing kernel Hilbert space $\mathcal{H}$ using a Gaussian RBF kernel. This term penalizes divergence between the fused feature distributions of clean-weather and adverse-weather samples, encouraging domain-invariant representations.

3.5 | Classification Head

The fused feature $\mathbf{g}^{\text{fused}}$ is passed to a two-layer MLP classification head:

$$\hat{y} = \text{softmax}(\mathbf{W}_2 \cdot \text{GELU}(\mathbf{W}_1 \mathbf{g}^{\text{fused}} + \mathbf{b}_1) + \mathbf{b}_2)$$

where $\mathbf{W}_1 \in \mathbb{R}^{d \times 2d}$, $\mathbf{W}_2 \in \mathbb{R}^{2d \times C}$, and GELU denotes the Gaussian Error Linear Unit activation.

3.6 | Training Objective

The total training loss is a weighted combination of the cross-entropy classification loss and the MMD domain alignment loss:

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{CE}}(\hat{y}, y) + \lambda \cdot \mathcal{L}_{\text{MMD}}$$

$$\mathcal{L}_{\text{CE}} = - \sum_{c=1}^C y_c \log \hat{y}_c$$

where λ is a balancing hyperparameter controlling the relative contribution of the domain alignment term. In experiments, $\lambda = 0.1$ is found to yield the best generalization on validation splits.

3.7 | Complexity Analysis

CDW-Net consists of approximately 62.4 million trainable parameters: the Swin-T backbone contributes $\sim 28.3\text{M}$ parameters, ResNet-50 $\sim 23.5\text{M}$, and the CDAF module and classification head $\sim 10.6\text{M}$. The inference complexity for an input of size 224×224 is approximately 9.8 GFLOPs, which is competitive with contemporary hybrid architectures used in weather recognition tasks [11, 10].

4 | Experimental Results

4.1 | Datasets

Multi-class Weather Dataset (MWD): A widely used benchmark for weather image classification containing images categorized across five weather conditions: sunny, cloudy, rainy, foggy, and snowy [11]. The dataset provides balanced class distribution with standard 70/15/15 train/validation/test splits.

Weather Phenomenon Database (WEAPD): A diverse benchmark of real-world weather condition images spanning multiple weather phenotypes including hail, lightning, rainbow, frost, and snow [11]. WEAPD presents a more challenging evaluation setting due to within-class visual variability and compound weather events.

Detection in Adverse Weather Nature (DAWN): A challenging dataset of 1,027 high-resolution traffic images captured under real-world adverse weather conditions (fog, rain, snow, and sandstorm) [15]. DAWN introduces a domain shift challenge, as images are collected under genuinely naturalistic conditions with

irregular degradation profiles. Table 1 summarizes the benchmark datasets used in this study, including the number of classes and total images available in each dataset.

Table 1. Summary of the Datasets Used in the Experiments.

Dataset	Classes	Total Images
MWD	4 (Sunrise, Shine, Rain, Cloudy)	1,125
WEAPD	11 Weather Phenomena	6,691
DAWN	4 (Fog, Snow, Rain, Sandstorm)	1,000

4.2 | Implementation Details

CDW-Net is implemented in PyTorch 2.0. The Swin Transformer branch is initialized with Swin-T weights pre-trained on ImageNet-1K; ResNet-50 is likewise initialized from ImageNet-1K pre-trained weights. The model is fine-tuned end-to-end for 100 epochs using the AdamW optimizer with an initial learning rate of 3×10^{-4} , weight decay 5×10^{-4} , and cosine annealing learning rate schedule. Mini-batch size is 32. Standard data augmentation includes random horizontal flipping, random resized cropping to 224×224 , color jitter (brightness ± 0.2 , contrast ± 0.2 , saturation ± 0.2), and random erasing. Experiments are conducted on an NVIDIA RTX 3060 Laptop GPU (6 GB VRAM) with 16 GB system RAM and Intel Core i7 12th Gen CPU. All experiments were conducted using a fixed random seed of 42. A preliminary sensitivity analysis indicated that λ values between 0.1 and 1.0 produced stable results, with $\lambda = X$ yielding the best validation accuracy.

4.3 | Comparison with State-of-the-Art

Table 2 reports the recognition accuracy on MWD, and CDW-Net achieves the top-1 accuracy of 97.84%, which surpasses the previous best method (ViT + Dual Attention [11]) by 0.37 percentage points with much fewer dedicated attention parameters thanks to the shared CDAF module. Similarly, the results on WEAPD and DAWN datasets are presented in Tables 3 and 4.

Table 2. Comparison on Multi-class Weather Dataset (MWD).

Method	Backbone	Accuracy (%)	F1-Score (%)	Params (M)
DenseNet-121 [13]	DenseNet-121	95.90	95.63	8.0
MobileNetV2+SVM [14]	MobileNetV2	93.84	93.52	3.4
InceptionV3 [19]	InceptionV3	93.81	93.44	23.8
MASK-CT [10]	CNN+Transformer	95.73	95.31	38.6
ViT + Dual Attn. [11]	ViT+CNN	97.47	97.39	86.4
CDW-Net (Ours)	Swin-T + ResNet50	97.84	97.71	62.4

Table 3. Comparison on WEAPD Dataset.

Method	Backbone	Accuracy (%)	F1-Score (%)
DenseNet-169 [19]	DenseNet-169	88.46	87.91
MASK-CT [10]	CNN+Transformer	89.22	88.84
ViT + Dual Attn. [11]	ViT+CNN	87.69	87.14
Truncated ResNet50 [20]	ResNet50	88.73	88.31
CDW-Net (Ours)	Swin-T + ResNet50	91.23	90.87

Table 4. Comparison on DAWN Dataset (Adverse Weather Recognition).

Method	Backbone	Accuracy (%)	Params (M)
YOLOv8 (detection-only) [21]	CSP-DarkNet	74.60	43.7
CNN Benchmark (ADS) [15]	Various CNN	81.35	—
DAGL-Faster (adapted) [16]	ResNet-101	83.14	68.5
CDW-Net (Ours)	Swin-T + ResNet50	89.67	62.4

In particular, CDW-Net produces the largest performance gain on the DAWN dataset, with a 6.53 percentage point increase in accuracy over the domain adaptation baseline, because the gain mainly comes from the explicit domain alignment enforced by the CDAF module, which is most beneficial when the training and test data distributions differ significantly.

4.4 | Ablation Study

To validate individual architectural components, Table 5 reports ablation results on the MWD validation set.

Table 5. Ablation Study on MWD Validation Set.

Configuration	Accuracy (%)	F1-Score (%)
CNN Branch Only (ResNet-50)	94.12	93.87
Transformer Branch Only (Swin-T)	95.63	95.41
Dual-Branch (Concat, no CDAF)	96.47	96.21
Dual-Branch + CDAF (no MMD)	97.18	97.02
CDW-Net (Full: Dual-Branch + CDAF + MMD)	97.84	97.71

The ablation results in Tab. 2 clearly demonstrate the effectiveness of each component of the proposed model, and using only the transformer branch is already better than the CNN branch alone by 1.51 percentage points, showing the importance of global attention for modeling the weather-related degradations. Using the proposed CDAF module can further improve the performance over naive feature concatenation by 0.71 percentage points, and using MMD-based domain alignment can bring an additional 0.66 percentage point improvement. All of these results confirm the importance of both the cross-branch attention fusion and explicit distribution alignment for robust cross-domain generalization.

5 | Discussion

5.1 | Interpretation of Results

The strong performance of CDW-Net on MWD, WEAPD, and DAWN also confirms that its dual-branch structure is capable of capturing the complementary characteristics of weather-induced degradation, because specifically, the transformer branch is more suitable to focus on spatially diffuse, large-scale patterns, which is beneficial to model the global visibility loss in foggy and hazy conditions. On the other hand, the convolutional branch captures some local edges and textures that are informative in rainy scenes, where the precipitation artifacts are more sparse and scatter over the image, while the CDAF module then adaptively reweights these two branches and dynamically emphasizes one over the other according to the content of each input image. Consequently, the adaptive fusion allows the network to adjust its recognition strategy according to the specific type of degradation, which leads to robust performance in diverse adverse weather scenarios.

CDAF's MMD-based domain alignment is particularly beneficial for DAWN, where test images are subject to natural, highly variable weather conditions that cannot be captured by any fixed training set, because by penalizing the divergence between the source and target feature distributions during training, the model

develops more domain-invariant representations that are less influenced by specific imaging conditions, thereby increasing its resilience to unseen degradation patterns.

The observation that CDW-Net performs much better than single-branch baselines is also consistent with the recent evidence in the literature [10, 11] that hybrid CNN–Transformer networks consistently outperform their unimodal counterparts on a variety of weather-related visual recognition tasks. Furthermore, the 3.47 pp improvement over the best CNN-only benchmark on DAWN underscores the practical significance of domain-adaptive training, consistent with findings reported by Yuan et al. [9] and Goel et al. [8] for domain-adaptive classification and detection.

5.2 | Failure Cases and Limitations

CDW-Net’s primary failure modes involve images with extreme, uniform degradation — particularly dense homogeneous fog at very low contrast — where both branches fail to extract discriminative features, and compound weather events (e.g., simultaneous rain and fog) that produce degradation patterns underrepresented in training data. Additionally, the model is evaluated on RGB-only inputs and does not leverage depth or LiDAR modality information, which is increasingly available in autonomous driving platforms [2]. The MMD alignment assumes that source and target feature distributions can be aligned in the RKHS induced by the RBF kernel, an assumption that may not hold for highly heterogeneous multi-weather target distributions.

Computational resource requirements — 62.4M parameters and 9.8 GFLOPs per inference — represent a constraint for edge deployment on embedded platforms, though the parameter count is substantially lower than competing methods such as ViT + Dual Attention [11] (86.4M parameters).

5.3 | Future Work

Several directions merit future investigation. First, incorporating a lightweight diffusion-based weather restoration module as a front-end pre-processor — following the approach of Özdenizci et al. [12] — may further improve performance on extremely degraded inputs. Second, extending CDW-Net to a multi-task framework that jointly predicts weather category and performs semantic segmentation would enhance its utility for autonomous driving perception. Third, exploring prompt-tuning of large pre-trained vision-language models for zero-shot or few-shot adverse weather recognition represents a promising direction, particularly for rare weather phenomena with limited training data [17]. Finally, incorporating contrastive degradation-aware pretraining — analogous to the approach used in MWFormer [4] — could further strengthen domain-invariant feature learning without requiring explicit domain labels.

6 | Conclusion

This paper proposed CDW-Net, a novel cross-domain hybrid vision architecture for adverse weather image recognition. By integrating a dual-branch encoder — combining Swin Transformer-based global attention with ResNet-50 local feature extraction — with a Cross-Domain Attention Fusion module incorporating MMD-based domain alignment, CDW-Net addresses the core limitations of existing approaches: their inability to generalize across multiple weather degradation types and their susceptibility to domain shift. Comprehensive evaluation on three benchmark datasets (MWD: 97.84%, WEAPD: 91.23%, DAWN: 89.67%) confirms that CDW-Net consistently outperforms state-of-the-art CNN and transformer baselines. Ablation studies validate the contribution of each architectural component, with the CDAF module and MMD alignment jointly contributing 1.37 pp improvement over a simple dual-branch concatenation baseline. CDW-Net provides a scalable and practical foundation for robust all-weather visual recognition in

safety-critical real-world systems, with clear future directions toward multi-task learning, edge deployment, and zero-shot generalization.

Funding

This research has no funding source.

Conflicts of Interest

The authors declare that there is no conflict of interest in the research.

Ethical Approval

This article does not contain any studies with human participants or animals performed by any of the authors.

References

- [1] Tahir, N. U. A., Zhang, Z., Asim, M., Chen, J., & ELAffendi, M. (2024). Object detection in autonomous vehicles under adverse weather: A review of traditional and deep learning approaches. *Algorithms*, 17(3), 103.
- [2] Eskandar, G., Marsden, R. A., Pandiyan, P., Döbler, M., Guirguis, K., & Yang, B. (2022, October). An unsupervised domain adaptive approach for multimodal 2d object detection in adverse weather conditions. In *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)* (pp. 10865-10872). IEEE.
- [3] Munir, A., Siddiqui, A. J., Anwar, S., El-Maleh, A., Khan, A. H., & Rehman, A. (2024). Impact of adverse weather and image distortions on vision-based UAV detection: A performance evaluation of deep learning models. *Drones*, 8(11), 638.
- [4] Zhu, R., Tu, Z., Liu, J., Bovik, A. C., & Fan, Y. (2024). Mwformer: Multi-weather image restoration using degradation-aware transformers. *IEEE Transactions on Image Processing*, 33, 6790-6805.
- [5] Song, Y., He, Z., Qian, H., & Du, X. (2023). Vision transformers for single image dehazing. *IEEE Transactions on Image Processing*, 32, 1927-1941.
- [6] Ali, A. M., Benjdira, B., Koubaa, A., El-Shafai, W., Khan, Z., & Boulila, W. (2023). Vision transformers in image restoration: A survey. *Sensors*, 23(5), 2385.
- [7] Valanarasu, J. M. J., Yasarla, R., & Patel, V. M. (2022). Transweather: Transformer-based restoration of images degraded by adverse weather conditions. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 2353-2363).
- [8] Goel, P., & Ganatra, A. (2023). Unsupervised domain adaptation for image classification and object detection using guided transfer learning approach and JS divergence. *Sensors*, 23(9), 4436.
- [9] Yuan, Z., Li, Y., & Yang, J. (2022, December). Improving hazy image recognition by unsupervised domain adaptation. In *2022 17th International Conference on Control, Automation, Robotics and Vision (ICARCV)* (pp. 311-316). IEEE.
- [10] Chen, S., Shu, T., Zhao, H., & Tang, Y. Y. (2023). MASK-CNN-Transformer for real-time multi-label weather recognition. *Knowledge-Based Systems*, 278, 110881.
- [11] Li, J., & Luo, X. (2023). A study of weather-image classification combining ViT and a dual enhanced-attention module. *Electronics*, 12(5), 1213.
- [12] Özdenizci, O., & Legenstein, R. (2023). Restoring vision in adverse weather conditions with patch-based denoising diffusion models. *IEEE transactions on pattern analysis and machine intelligence*, 45(8), 10346-10357.
- [13] Albelwi, S. A. (2022). Deep architecture based on DenseNet-121 model for weather image recognition. *International Journal of Advanced Computer Science and Applications*, 13(10).
- [14] Yildirim, M., Çinar, A., & Cengil, E. (2022). Classification of the weather images with the proposed hybrid model using deep learning, SVM classifier, and mRMR feature selection methods. *Geocarto International*, 37(9), 2735-2745.
- [15] Afxentiou, V., & Vladimirova, T. (2025). Evaluation of CNN-based approaches to adverse weather image classification for autonomous driving systems. *IEEE Open Journal of Intelligent Transportation Systems*, 6, 204-229.
- [16] Hu, M., Wu, Y., Yang, Y., Fan, J., & Jing, B. (2023). DAGL-Faster: Domain adaptive faster r-cnn for vehicle object detection in rainy and foggy weather conditions. *Displays*, 79, 102484.
- [17] Hussein, M. K., Alqassab, A., & Alkahla, L. T. (2025). Image Classification Using Deep Learning: A Systematic Review. *Neutrosophic Optimization and Intelligent Systems*, 8, 33-40. <https://doi.org/10.61356/j.nois.2025.8646>
- [18] Saeed, J. N., & Hussein, M. K. (2025). A Multi-ViT's-Based Approach for Automatic Rice Leaf Disease Classification. *Iraqi Journal of Science*, 3938-3950.

- [19] Wulandari, V., Sari, W. J., Al-Sawaff, Z. H., & Manickam, S. (2025). Comparative analysis of weather image classification using CNN algorithm with InceptionV3, DenseNet169 and NASNetMobile architecture models. *Public Research Journal of Engineering, Data Technology and Computer Science*, 2(2), 81-92.
- [20] Ouattara, H., Salmane, P. H., Duthon, P., Bernardin, F., & Aider, O. A. (2026). Style-Based Neural Architectures for Real-Time Weather Classification. *arXiv preprint arXiv:2604.18251*.
- [21] Aloufi, N., Alnori, A., & Basuhail, A. (2024). Enhancing autonomous vehicle perception in adverse weather: A multi objectives model for integrated weather classification and object detection. *Electronics*, 13(15), 3063.

Disclaimer/Publisher's Note: The perspectives, opinions, and data shared in all publications are the sole responsibility of the individual authors and contributors, and do not necessarily reflect the views of Sciences Force or the editorial team. Sciences Force and the editorial team disclaim any liability for potential harm to individuals or property resulting from the ideas, methods, instructions, or products referenced in the content.