electronics

MDPI

Article

PF-ConvNeXt: An Adverse Weather Recognition Network for Autonomous Driving Scenes

Quanxiang Wang, Zhaofa Zhou * and Zhili Zhang

School of Missile Engineering, Rocket Force University of Engineering, Xi'an 710025, China; wqx15609278062@163.com (Q.W.); zzl202@hhu.edu.cn (Z.Z.)
* Correspondence: zzftxy@163.com

Abstract

Rain, snow, fog, and dust can degrade road-scene images, blur fine details, and consequently reduce the reliability of perception systems for autonomous driving. To address this problem, this paper proposes PF-ConvNeXt, an adverse weather recognition model built upon the ConvNeXt architecture. First, a lightweight pyramid split attention (PSA) module is introduced to enable multi-scale feature fusion, so that both global degradation patterns and local texture details can be captured simultaneously. Second, a feature enhancement channel and spatial attention module (FECS) is designed. It adaptively recalibrates features along the channel and spatial paths, thereby suppressing interference from complex backgrounds and noise. Third, during training, Focal Loss is adopted to strengthen learning for hard samples and minority weather categories, alleviating recognition bias caused by class imbalance. Experiments are conducted on a dataset of 5000 images constructed by integrating RTTS, DAWN, and a self-collected rainy-weather dataset. The results show that PF-ConvNeXt achieves 90.16% accuracy, 95.24% mean average precision, and a 92.18% F1-score. It outperforms the ConvNeXt baseline by 4.74%, 5.46%, and 5.95%, respectively, and surpasses multiple mainstream classification models. This study provides an effective recognition framework for robust environmental perception under challenging weather conditions and demonstrates promising potential for practical deployment.

Keywords: adverse weather; ConvNeXt; multi-scale feature fusion; Focal Loss

1. Introduction

With advances in artificial intelligence, computer vision has been widely adopted in autonomous driving. In real-world operation, adverse conditions such as rain, snow, fog, and dust reduce image contrast, blur fine details, and weaken object boundaries, which degrades downstream detection and tracking. Therefore, accurately recognizing the current weather condition at the perception front end and providing reliable cues to subsequent modules is essential for improving the robustness of autonomous driving systems.

Convolutional neural networks (CNNs) have been widely used in deep learning because of their strong feature representation capability. Since LeCun et al. [1] developed LeNet for handwritten digit recognition, a series of classic architectures, including AlexNet [2], VGGNet [3], and GoogLeNet [4], have continuously advanced image classification. In recent years, Transformers have shown remarkable performance in vision tasks [5]; however, their computational overhead and deployment cost remain limiting factors for certain real-time systems. Liu et al. [6] proposed ConvNeXt, which revisits and refines the design of convolutional networks. This enables pure convolutional architectures

Received: 21 January 2026
Revised: 19 February 2026
Accepted: 23 February 2026
Published: 24 February 2026

to compete with mainstream vision Transformers in accuracy while retaining the efficiency advantages of convolutions, thereby offering a more practical option for deployment.

Nevertheless, adverse-weather recognition remains challenging. First, weather phenomena are stochastic and involve complex imaging processes [7]. Second, even within the same weather category, appearance can vary substantially with intensity, viewpoint, illumination, and scene context [8]. Third, public adverse-weather datasets are often limited in size and imbalanced across categories, which restricts model generalization [9]. To address these issues, this paper develops an adverse-weather recognition model for real-world road scenes based on ConvNeXt. The proposed method aims to improve recognition accuracy and cross-scene stability, providing reliable prior information for downstream perception and decision-making under adverse conditions.

Overall, the purpose of the present paper is as follows:

(1) PF-ConvNeXt for adverse-weather recognition. Using ConvNeXt as the backbone, we build an enhanced network that better captures degradation patterns caused by rain, snow, fog, and dust, enabling stable multi-class recognition.
(2) We propose feature enhancement channel and spatial attention module (FECS) to adaptively recalibrate features through two complementary paths: channel semantics and spatial locations. This emphasizes weather-related responses while suppressing interference from complex backgrounds and noise, thereby improving cross-scene robustness.
(3) We introduce a lightweight pyramid split attention (PSA) module to perform multi-scale feature fusion. Through multi-scale splitting and cross-scale interaction, the module captures both global haze-like scattering patterns and fine-grained cues such as rain streaks and snow textures, enhancing the representation of weather cues across different intensities and scales. In addition, Focal Loss is adopted to emphasize hard samples and minority classes, alleviating class-imbalance issues.
(4) We build a dataset by integrating RTTS, DAWN, and a self-collected rainy-weather dataset, and expand it to 5000 samples via augmentation. Extensive ablation and comparative experiments are conducted to validate the effectiveness of each component. Results demonstrate that the proposed method outperforms multiple mainstream models in terms of accuracy and precision.

2. Related Work

2.1. CNN-Based Image Classification Methods

CNNs have long dominated image classification due to local receptive fields, weight sharing, and hierarchical feature extraction. Krizhevsky et al. introduced AlexNet, one of the earliest deep CNNs for large-scale image classification, and improved feature learning through stacked convolution and pooling operations. Simonyan and Zisserman proposed VGGNet, which increases representational capacity by stacking small convolution kernels; however, its computational cost is high for resource-constrained settings. Szegedy et al. introduced GoogLeNet with the Inception module to enable multi-scale convolution and pooling within a unified design. He et al. [10] proposed ResNet with residual connections to ease optimization of very deep networks. Huang et al. [11] introduced DenseNet, which connects each layer to all previous layers to encourage feature reuse and improve information flow. More recently, Liu et al. proposed ConvNeXt, which modernizes CNNs design by incorporating training and architectural practices popularized by vision Transformers while retaining a pure convolutional backbone. Shigesawa et al. [12] proposed a method to address the problem of low accuracy caused by illumination changes in winter road surface classification using on-vehicle cameras in snowy regions. This method standardizes images into Core-Day and Core-Night styles via CycleGAN, extracts feature with

MobileNet [13]. Chung [14] proposed a lightweight framework that integrates coordinate attention (CA) with L1-regularized channel pruning, achieving a significant reduction in the parameters and computational complexity of CNN-based image classification models while maintaining or improving classification accuracy. In the context of autonomous driving under adverse weather, Manivannan et al. [15] proposed a multi-level knowledge distillation framework to improve weather classification accuracy while reducing computational overhead, which is beneficial for onboard deployment. Moreover, Introvigne et al. [16] introduced RECNet for real-time environment condition classification from a single RGB frame and validated its effectiveness on a newly defined taxonomy and dataset pipeline, achieving real-time performance.

Despite their success, CNNs rely primarily on local convolution operations, which limits the effective receptive field and makes global context modeling less direct. In complex scenes where long-range dependencies are important, additional mechanisms or architectural refinements are often required to improve performance.

2.2. Transformer-Based Image Classification Methods

Following the success of Transformers in natural language processing, self-attention has been adopted for vision tasks, leading to Transformer-based image classification. Dosovitskiy et al. [17] proposed the Vision Transformer (ViT), which splits an image into fixed-size patches and models global relationships using self-attention. While ViT can capture long-range dependencies effectively, it lacks some of the local inductive biases of CNNs and typically benefits from large-scale training data. To improve data efficiency, Touvron et al. [18] proposed Data-efficient Image Transformer (DeiT). This method works well in enhancing the generalization of smaller data, and it is best applied in the scenario of the limited data. Liu et al. [19] introduced Swin Transformer with shifted windows to reduce the cost of global self-attention while enabling cross-window interaction. Cui et al. [20] proposed an improved Swin Transformer model, integrating an efficient channel attention (ECA) module after its self-attention mechanism to dynamically adjust the weights of channel features, thereby enhancing detection accuracy. Wang et al. [21] proposed Pyramid Vision Transformer (PVT), which builds hierarchical multi-scale representations to support dense prediction tasks; however, Transformer-based backbones can still incur relatively high computation and memory costs, especially for high-resolution inputs. For weather recognition, Chen et al. [22] proposed MASK-CNN-Transformer (MASK-CT) for real-time multi-label weather recognition by combining CNN feature extraction with a Transformer encoder and a masking strategy to improve generalization.

Although Transformer-based approaches are effective at modeling long-range dependencies and complex visual patterns, their computation and memory requirements can be prohibitive for real-time, resource-constrained onboard systems. Achieving high accuracy while maintaining low latency and low power consumption remains an active challenge for practical deployment.

2.3. CNN-Transformer Hybrid Methods for Image Classification

To combine the local modeling strengths of CNNs with the global modeling capability of Transformers, various CNN-Transformer hybrid architectures have been proposed. Wu et al. [23] introduced CvT, which uses convolutional token embedding to strengthen local inductive bias within a Transformer framework, but its computational complexity can be high. Dai et al. [24] proposed CoAtNet, which interleaves convolutional layers and attention layers to capture both local and global information; however, its architectural complexity increases computational and memory demands. Mehta [25] proposed MobileViT, which integrates lightweight convolutions with Transformer modules to improve

efficiency and is suitable for mobile and embedded scenarios, although performance may drop on highly complex visual tasks. Recent studies also explore multi-task or hybrid designs for adverse-weather understanding in driving scenes. Aloufi et al. [26] proposed a multi-objectives framework that integrates weather classification and object detection under adverse weather using camera data and evaluated it on the DAWN dataset. From the data perspective, WEDGE [27] provides a multi-weather autonomous driving dataset built with generative vision-language models, offering broader weather coverage and benchmarks for both weather classification and detection.

In summary, CNN-Transformer hybrid designs are promising for multi-scale feature modeling and complex-scene understanding, but they often introduce additional parameters and computation. Balancing performance and efficiency remain important for high-resolution processing and real-time applications.

To conclude, CNN-based, Transformer-based, and hybrid CNN-Transformer image classification approaches possess their advantages in terms of the feature modeling potential. In order to solve the multiple-scale feature complexity, experimental background interference, or class imbalance problems in adverse weather recognition, the proposed paper suggests PF-ConvNeXt that is based on the ConvNeXt architecture. PF-ConvNeXt proposes multi-scale feature fusion, spatial-channel joint attention, and hard-sample aware loss function to result in strong adverse weather image classification.

3. Method

3.1. ConvNeXt

ConvNeXt revisits convolutional network design by incorporating several architectural and training practices inspired by vision Transformers, while retaining a pure convolutional backbone. It follows a hierarchical stage-wise structure and is built around residual blocks that combine large-kernel depthwise convolutions for spatial modeling and 1×1 pointwise convolutions for channel mixing. A typical ConvNeXt block applies a large-kernel depthwise convolution, followed by LayerNorm and two pointwise 1×1 convolutions with a non-linear activation between them. The inverted bottleneck design expands channels for feature transformation and then projects back to the original dimension, providing a favorable accuracy-efficiency trade-off. ConvNeXt also uses independent downsampling layers between stages to stabilize optimization and improve performance without substantially increasing computational complexity.

3.2. PF-ConvNeXt

Adverse-weather recognition remains challenging because weather effects are stochastic and vary significantly with intensity, viewpoint, illumination, and scene context. Although ConvNeXt performs well on general image recognition tasks, there is room for improvement on adverse-weather classification. To address these limitations, we propose PF-ConvNeXt, an enhanced network built on ConvNeXt. PF-ConvNeXt integrates multi-scale feature fusion and a feature-enhanced spatial-channel attention module to improve robustness under diverse adverse-weather conditions. Figure 1 illustrates the overall architecture of PF-ConvNeXt.

The input to the proposed network is an RGB road-scene image of size $224 \times 224 \times 3$. A 4×4 convolutional stem layer with stride 4 is applied to perform initial feature embedding and spatial downsampling. This reduces computational cost while preserving important texture and edge information, providing a solid representation for subsequent feature learning.

First, to improve the model's ability to capture salient structures under rain, snow, and fog, we introduce a lightweight PSA mechanism at early stages to build a multi-scale feature

fusion module. By splitting features across scales and enabling cross-scale interaction, PSA captures complementary information from different receptive fields and adaptively fuses global and local cues. This helps reduce background interference and improves discrimination of weather-related patterns under complex imaging noise.

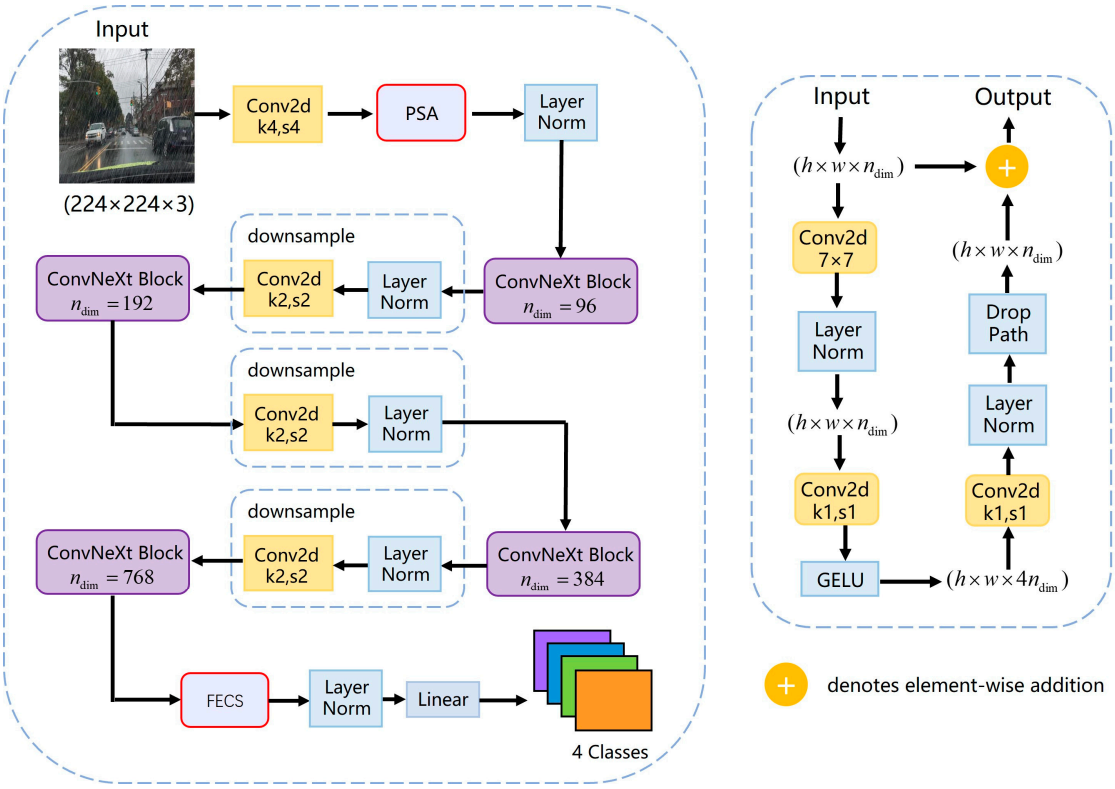

Figure 1. The PF-ConvNeXt network architecture (**left**) and the ConvNeXt block module (**right**).

Second, at higher semantic stages, we insert the FECS module. FECS recalibrates features along two paths: a spatial branch that emphasizes regions directly related to weather phenomena and a channel branch that strengthens key semantic responses. By jointly enhancing spatial and channel features, the model better distinguishes weather-induced degradation from inherent scene textures, improving stable recognition under diverse scenes and high interference.

Third, at the classification head, global average pooling aggregates the final feature maps, followed by normalization and a linear layer to output predictions for each weather category. To address class imbalance and the presence of easy vs. hard samples in real-world data, we adopt Focal Loss during training. By down-weighting easy samples and emphasizing hard samples, Focal Loss mitigates bias from imbalanced class distributions and improves recognition of less frequent and more challenging weather categories.

3.3. PSA

To enhance the network's ability to represent multi-scale weather cues at a finer granularity, this paper introduces a lightweight PSA mechanism [28], as shown in Figure 2. Adverse weather phenomena in images often exhibit significant scale differences. For example, fog-induced low-contrast scattering across a large area is a global feature, while rain streaks, snowflakes, and dust particles manifest as local detailed textures. Additionally, nighttime glare and dark region structures lead to changes in brightness distribution at different scales. Therefore, feature extraction with a single receptive field struggles to simultaneously capture both global structures and local details. PSA enhances the model's

ability to capture complex weather patterns through multi-scale splitting, feature fusion, and attention recalibration, all with relatively low computational overhead.

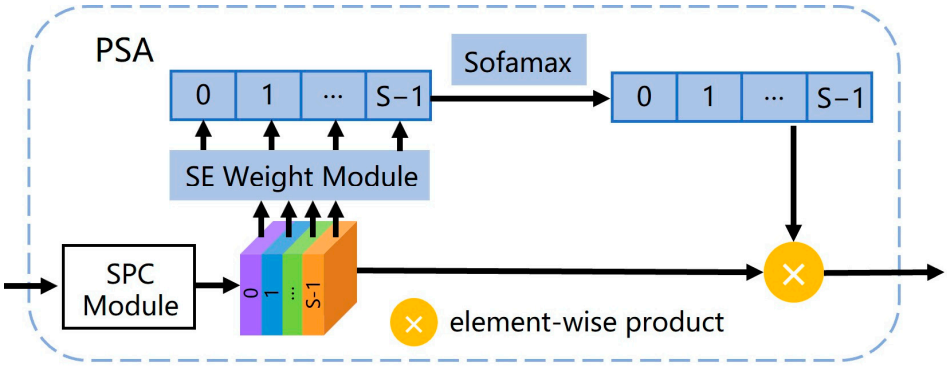

Figure 2. PSA structure diagram.

The processing flow of the PSA module is as follows: First, as shown in Figure 3, the feature map X output from the convolutional layer is input into the SPC module, which divides it into S sub-feature maps $\{X_1, X_2, \ldots, X_S\}$ along the channel dimension. Then, different-sized convolution kernels are applied to each sub-feature map for feature extraction, allowing the model to obtain multi-scale features under different receptive fields.

$$F_S = Conv(X_S, k),\ k \in \{k_1, k_2, k_3, k_4\} \tag{1}$$

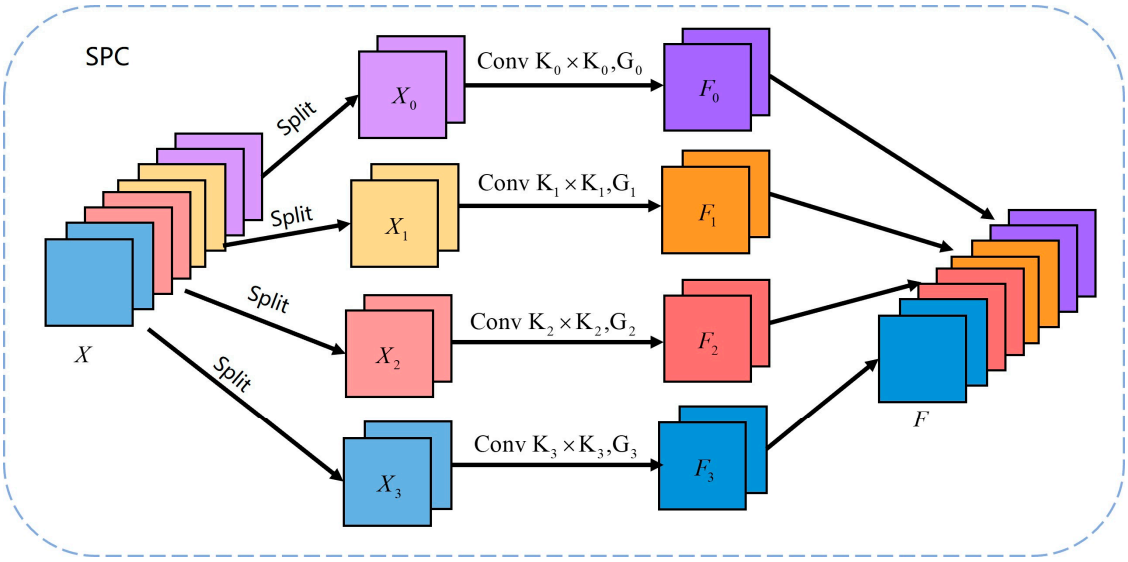

Figure 3. SPC structure diagram.

The resulting feature maps $\{F_1, F_2, \ldots, F_S\}$ can focus on weather-related information at different scales, such as large-scale fog coverage, and small-scale rain streaks and snowflake textures. Since the feature maps from different scales may have varying sizes, this paper uses bilinear interpolation to uniformly scale them to the same resolution, obtaining the standardized features.

$$F_i' = Resize(F_i, n_t), \forall i \tag{2}$$

Next, the features from each scale are aggregated and fused to form a comprehensive feature map F.

$$F = \sum_{i=1}^{S} F_i' \tag{3}$$

Building on multi-scale fusion, to further highlight key weather cues and suppress complex background noise, such as road textures and building edges, this paper introduces an attention weight vector A for adaptive recalibration of the fused features. Specifically, global average pooling is first applied to F, and then a fully connected mapping is used to obtain the channel attention.

$$A = \sigma(W_1 f_{GAP}(F)) \tag{4}$$

Here, $f_{\mathrm{GAP}}(\cdot)$ represents global average pooling, $\sigma(\cdot)$ is the activation function, and W_1 denotes the learnable parameters. The attention weights are then applied to the fused features, resulting in the enhanced feature map.

$$F_a = F \otimes A \tag{5}$$

Finally, to retain the original input features and enhance the stability of gradient propagation, a residual connection is used to add F_a and the input feature map X element-wise, resulting in an output with richer semantics.

$$F_{final} = F_a + X \tag{6}$$

In summary, the PSA module effectively integrates multi-scale information with a lightweight design and strengthens the discriminative features related to adverse weather through the attention mechanism, while suppressing background interference. Embedding this module into the ConvNeXt-based adverse weather recognition network helps enhance the model's accuracy and robustness under varying weather intensities, different scene backgrounds, and noisy conditions, while maintaining good computational efficiency. This makes it suitable for deployment in resource-constrained or real-time applications.

3.4. FECS

To improve the quality of discriminative features under complex imaging conditions such as rain, snow, fog, dust, and nighttime, this paper designs the FECS module, as shown in Figure 4. Adverse weather introduces sparse noise like rain streaks and snowflakes, low contrast and detail loss due to fog scattering, and glare and dark noise at night. These disturbances often lead to abnormal peaks and local pseudo-features in the feature responses. FECS enhances the effective weather-related cues and suppresses background and noise interference through recalibration in the channel dimension and multi-scale enhancement in the spatial dimension, thereby improving the stability and generalization of adverse weather classification with low additional computational cost.

In the channel attention branch, FECS simultaneously performs global average pooling, global max pooling, and global median pooling on the input feature map to obtain three complementary global statistical descriptions. Among these, median pooling is less sensitive to outliers and effectively reduces the impact of noise peaks such as highlighted raindrops or reflected glare on the channel weight estimation. The results of the three pooling operations are mapped into attention vectors via a shared MLP, which are then fused to produce the final channel attention map, used to apply channel-wise weighting to the input features.

$$F_c = \sum_{p \in \{\mathrm{Avg,Max,Med}\}} \sigma\big(\mathrm{MLP}\big(\mathrm{Pool}_p(F)\big)\big) \tag{7}$$

$$F' = F_c \odot F \tag{8}$$

This design allows the network to more reliably emphasize channel responses related to adverse weather, such as the low-frequency structures of fog, rain, and snow texture features, while suppressing channel activations dominated by background textures and noise.

In the spatial attention branch, FECS uses multi-scale depthwise separable convolutions to model the spatial relationships of F'. First, a 5×5 depthwise convolution is used to extract the basic spatial features. Then, depthwise convolutions at different scales are introduced in parallel to capture weather patterns with significant directional and scale differences. After element-wise fusion of the outputs at each scale, a 1×1 convolution is applied to generate the spatial attention map, which is then used to weight the features.

$$F_s = \sum_{i=1}^{n} D_i(F') \tag{9}$$

$$F'' = \mathrm{Conv}_{1\times1}(F_s) \odot F' \tag{10}$$

The multi-scale spatial attention effectively highlights the areas or structural distributions that are truly weather-related in complex backgrounds, reducing the interference of irrelevant information such as road textures and building edges on classification decisions. Overall, FECS enhances feature representation by improving channel selection robustness and providing comprehensive spatial enhancement.

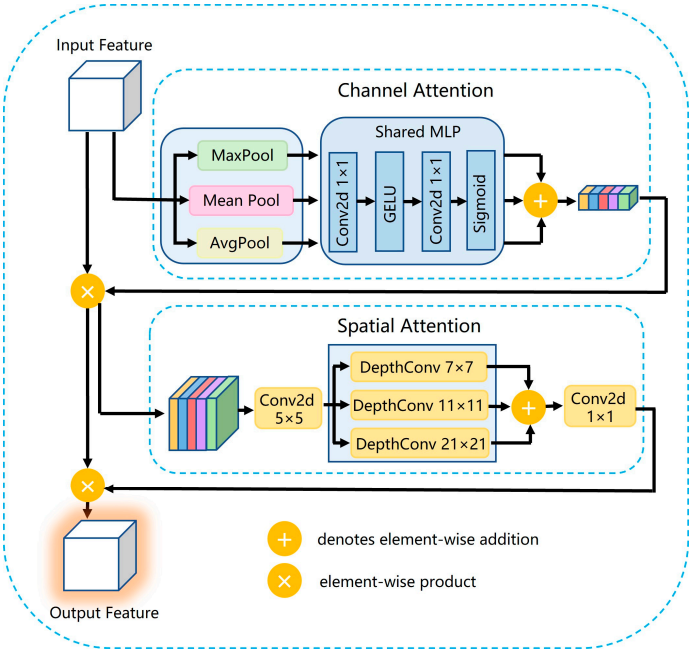

Figure 4. FECS structure diagram.

3.5. Focal Loss

Currently, most image classification models typically use cross-entropy loss as the optimization objective. This loss works well when the class distribution is relatively balanced. However, in real-world adverse weather datasets, issues such as class imbalance and a high proportion of easy-to-classify samples are common. For example, sunny day samples far outnumber samples from rare weather categories like dense fog, heavy snow, and sandstorms. Additionally, the abundance of clear samples often overshadows the gradient contribution of difficult samples. In such scenarios, cross-entropy loss is dominated by the majority of easy-to-classify samples, leading to insufficient learning of minority categories and boundary samples, which results in recognition bias and reduced generalization ability.

To enhance the model's focus on hard-to-classify samples and minority weather categories, this paper introduces Focal Loss [29] to improve the training objective. Its basic form is shown in the following equation:

$$FL(p_t) = -(1 - p_t)^{\gamma} \ln(p_t) \tag{11}$$

In this equation, p_t represents the predicted probability of the true class by the model, where a higher value indicates that the sample is easier to classify correctly. $(1 - p_t)^{\gamma}$ is the modulation factor, used to dynamically weight samples of varying difficulty. γ is the focusing parameter, which controls the degree to which the loss emphasizes difficult samples. As γ increases, the weight of easy-to-classify samples (with higher p_t) is suppressed, while the loss contribution of hard-to-classify samples (with lower p_t) increases. According to the sensitivity study in Section 4.5.1, we choose a focusing parameter γ of 2 for all experiments. This encourages the model to focus more on optimizing difficult samples and minority categories during training, effectively alleviating the training bias caused by imbalanced sample distribution.

4. Experiment

4.1. Experimental Design

Experiments were conducted on a Windows 10 (64-bit) workstation equipped with an NVIDIA GeForce RTX 4060 GPU (8 GB). All models were trained and evaluated under the same configuration. Training ran for 100 epochs with a batch size of 4. The initial learning rate was set to 0.001, and the Adam optimizer was used. Weight decay and early stopping were applied to reduce overfitting and improve generalization. Table 1 summarizes the hardware and software configuration.

Table 1. Experimental Hardware and Software Configuration.

Software/Hardware	Version
Operating System	Windows 10
CPU	Intel(R) Core(TM) i7-13650HX
GPU	NVIDIA GeForce RTX 4060
Memory	24 GB
Programming Language	Python 3.10
Deep Learning Framework	PyTorch 2.0.1
Parallel Computing Platform	CUDA 11.8

4.2. Dataset

The dataset used in this study consists of three parts. The first part is the RTTS dataset [30], from which 500 fog images were selected. The second part is the DAWN dataset [31], which provides 300 sandstorm images and 200 snow images. The third part is a self-collected rainy-weather dataset containing 500 images. To better approximate real-world conditions—where recognition is affected by illumination variations, sensor noise, and partial occlusions—and to improve generalization, we applied data augmentation, including random brightness adjustment, random Gaussian noise, random occlusion, and random rotation. Note that random occlusion is not intended to strictly model the physical degradation caused by adverse weather; instead, it is used to mimic practical disturbances such as partial occlusions and camera contamination, thereby improving robustness to local information loss. Specifically, the brightness factor was randomly sampled from [0.6, 1.4]; Gaussian noise with zero mean was added with a standard deviation randomly sampled from [0.01, 0.05] (images normalized to [0, 1]); random occlusion was implemented by masking 1–3 randomly placed rectangular patches, each covering 2–10% of the image area; and random rotation was applied with an angle uniformly sampled from $[-10°, +10°]$. Starting from 1500 raw images, we expanded the dataset to 5000 images by proportional augmentation sampling rather than applying all four augmentations to every image. Specifically,

for each class, we repeatedly generated augmented samples by randomly applying one augmentation to a randomly selected raw image until the target number of samples for that class was reached, while maintaining the original class ratio. The final class distribution is summarized in Table 2 (Fog: 1667; Rain: 1667; Sandstorm: 1000; Snow: 666). The dataset was split into training and test sets with an 8:2 ratio, and the per-class numbers are also reported in Table 2. Example images and augmented samples are shown in Figures 5 and 6.

Table 2. Dataset statistics and split.

Class	Raw Images	After Augmentation	Train (80%)	Test (20%)
Fog (RTTS)	500	1667	1333	334
Rain (Self-collected)	500	1667	1334	333
Sandstorm (DAWN)	300	1000	800	200
Snow (DAWN)	200	666	533	133
Total	1500	5000	4000	1000

Figure 5. Dataset examples.

Figure 6. Data augmentation image examples.

4.3. Evaluation Metrics

To assess the recognition performance of the proposed model, three metrics are used: accuracy (A), F1-score (F_1), and mean average precision (f_{mAP}). Accuracy reflects the overall proportion of correctly predicted samples; however, under class imbalance, accuracy alone may obscure performance on minority classes. F1-score is the harmonic mean of precision and recall. f_{mAP} summarizes the average performance across all classes. The definitions of these metrics are as follows:

$$A = \frac{TP + FN}{TP + TN + FP + FN} \times 100\% \tag{12}$$

$$F_1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \times 100\% \tag{13}$$

$$f_{\mathrm{mAP}} = \frac{1}{m} \sum_{i=1}^{m} \mathrm{AP}_i \tag{14}$$

In the formula, TP is the number of samples predicted as positive and correctly classified as positive by the model. FP: The number of samples predicted as positive but actually negative. TN is the number of samples predicted as negative and correctly

classified as negative. FN is the number of samples predicted as negative but actually positive. $AP = \int_0^1 P(R)dR$ represents the average precision for a single weather category, and m denotes the total number of categories. The higher the AP value, the better the model's performance.

The confusion matrix provides an intuitive way to show the correspondence between the model's predicted classes and the true labels, making it a commonly used tool for evaluating the performance of classification models. In the confusion matrix, the horizontal axis typically represents the predicted classes, and the vertical axis represents the true classes. When the values in the matrix are primarily concentrated along the main diagonal and the diagonal elements are large, it indicates that the model's classification performance is more accurate and the overall classification effect is better. Based on this, this paper uses the confusion matrix for visualizing the model's classification performance and further identifies which categories are more likely to be confused during the prediction process. This allows for targeted error analysis and improvements.

4.4. Ablation Analysis

To validate the individual contributions and synergistic effects of the introduced PSA, Focal Loss, and the designed FECS modules in the adverse weather recognition task, systematic ablation experiments were conducted under identical training strategies, data splits, and experimental environments. ConvNeXt was used as the baseline model, and on this basis, individual modules and different module combinations were progressively introduced. The performance of the models was compared using A, f_{mAP}, F_1, Params and GFLOPs as evaluation metrics. The experimental results are shown in Table 3, where "√" indicates that the corresponding module was introduced.

Table 3. Ablation experiments.

Model	PSA	FECS	Focal Loss	A (%)	f_{mAP} (%)	F_1 (%)	Params (M)	GFLOPs
ConvNeXt				85.42	89.78	86.23	28.0	4.5
	√			86.84	91.02	88.12	28.3	4.6
		√		87.32	91.66	88.69	28.5	4.8
			√	86.23	90.45	87.35	28.0	4.5
	√	√		88.17	92.96	89.67	28.9	4.9
	√		√	89.23	94.05	90.86	28.3	4.6
	√	√	√	90.16	95.24	92.18	30.1	5.0

In the single-module ablation experiments, each improvement strategy has had a positive impact on the model's performance. First, after introducing the PSA module, the model's accuracy increased by 1.42% compared to the baseline model, with parameter count and computational load increasing by only 0.3 M and 0.1 GFLOPs, respectively. This improvement is mainly attributed to PSA's effective integration of multi-scale features, enabling the network to simultaneously model global fog, low-contrast information, and local rain streaks, snow, and other fine-grained textures across different receptive fields. This enhances the model's stability in representing complex adverse weather conditions. Second, after introducing the FECS module, the model's accuracy f_{mAP} and F_1 increased by 1.90%, 1.88%, and 2.46%, respectively, accompanied by an increase of 0.5 M in parameters and 0.3 GFLOPs in computation. This demonstrates that FECS, through channel recalibration and spatial enhancement mechanisms, highlights key feature responses related to weather phenomena such as rain, snow, fog, and sandstorms, while suppressing background textures and noise interference, thus improving the stability of classification discrimination. Finally, after replacing the conventional loss function with Focal Loss, the model's focus on

hard-to-classify samples and minority weather categories increased. The model's accuracy improved by 0.81% over the baseline without introducing any additional parameters or computational cost. This reflects the loss function's effectiveness in enhancing the model's ability to discriminate difficult samples at the objective-optimization level.

Clear complementarity is observed in the multi-module combination experiments. Adding PSA and FECS improves accuracy by 2.75% over the baseline, with parameter and computational increases of 0.9 M and 0.4 GFLOPs, respectively. This indicating that multi-scale representations from PSA provide stronger support for the attention refinement in FECS by balancing global structure and local details. When PSA is combined with Focal Loss, accuracy improves by 3.81% over the baseline, suggesting synergy between multi-scale feature modeling and hard-sample reweighting. When PSA, FECS, and Focal Loss are used together, the model achieves the best performance, with accuracy improving by 4.74% over the baseline, and the other metrics reaching their highest values.

In conclusion, the ablation study systematically verified the effectiveness of the proposed PSA, FECS, and Focal Loss in the adverse weather recognition task. The single-module experiments show that each improvement strategy enhances the model's feature modeling and classification ability from different aspects. PSA strengthens multi-scale representation capability, FECS enhances the focus on key weather features, and Focal Loss effectively alleviates training bias caused by class imbalance and hard sample recognition. When PSA, FECS, and Focal Loss are all incorporated, the model achieves the best performance: accuracy improves by 4.74% over the baseline, and the other metrics also reach their highest values. The complete model then contains 30.1 M parameters and requires 5.0 GFLOPs, which remains within a reasonable range compared to the baseline. This indicates that the proposed improvements substantially boost recognition performance while maintaining favorable computational deployability.

4.5. Results of Comparative Experiments

4.5.1. Sensitivity Analysis of γ in Focal Loss

The focusing parameter γ in Focal Loss governs the decay rate of sample weights based on classification difficulty. To investigate its influence on model performance, we fixed the PF-ConvNeXt architecture and all other training hyperparameters, varying only the value of γ. The experimental results are presented in Table 4. As shown in Table 4, with standard cross-entropy (corresponding to a γ value of 0), the model achieves an accuracy of 84.80%, f_{mAP} of 88.90%, and F_1 of 86.00%. When Focal Loss is introduced with γ set to 1, all three metrics improve notably, indicating that Focal Loss effectively alleviates class imbalance. The model attains its best performance at γ equal to 2, with accuracy of 86.23%, f_{mAP} of 90.45%, and F_1 of 87.35%, which represent improvements of 1.43%, 1.55%, and 1.35% over the baseline, respectively. As γ is further increased to 3 and 4, performance gradually declines, suggesting that excessive focusing suppresses easy samples too aggressively and thus impairs feature learning. In summary, a γ value of 2 proves to be the optimal and robust choice for this task; therefore, all experiments in this study adopt this setting.

Table 4. Sensitivity of γ in Focal Loss.

γ	A (%)	f_{mAP} (%)	F_1 (%)
0	84.80	88.90	86.01
1	85.60	89.80	86.80
2	86.23	90.45	87.35
3	85.90	90.00	87.04
4	85.52	89.56	86.45

4.5.2. Experimental Comparison with Other Models

To further validate the effectiveness of the proposed PF-ConvNeXt model in the adverse weather recognition task, this study selects nine representative image classification models for comparison, including ConvNeXt, ResNet50, MobileNet, MobileViT, FasterNet [32], QKFormer [33], TLENet [34], GAC-SNN [35], and ATONet [36]. All experiments were conducted under the same dataset partition, training strategy, and evaluation metrics to ensure a fair comparison. The detailed performance comparison results are shown in Table 5.

Table 5. Comparative experiments.

Model	A (%)	f_{mAP} (%)	F_1 (%)	Params (M)	GFLOPs	FPS
ConvNeXt	85.42	89.78	86.23	28.0	4.5	90
ResNet50	86.59	91.06	87.67	25.5	4.1	115
MobileNet	87.03	92.94	88.35	4.2	1.1	96
MobileViT	86.12	91.64	87.23	5.6	1.8	56
FasterNet	86.78	92.53	87.96	31.1	4.5	92
QKFormer	89.82	95.04	91.86	16.5	3.0	33
TLENet	88.24	94.12	90.37	22.0	3.8	24
GAC-SNN	89.11	94.65	91.24	18.0	3.3	28
ATONet	87.42	93.13	89.14	12.3	2.2	40
PF-ConvNeXt	90.16	95.24	92.18	30.1	5.0	78

The experimental results show that, among the selected baseline models, MobileNet achieves relatively strong performance, with accuracy, f_{mAP}, and F_1 reaching 87.03%, 92.94%, and 88.35%, respectively, while maintaining the lowest parameter count and computational cost, highlighting its high efficiency. FasterNet and ResNet50 follow closely, with accuracies of 86.78% and 86.59%. Several recently proposed models, such as QKFormer and GAC-SNN, also demonstrate competitive results. Several recently proposed models also show competitive performance with well-balanced cost. QKFormer achieves 95.04% f_{mAP} and an F_1 of 91.86%, using 16.5 M parameters and 3.0 GFLOPs, while GAC-SNN attains 89.11% accuracy with 18.0 M parameters and 3.3 GFLOPs.

Nevertheless, the proposed PF-ConvNeXt delivers the best performance across all three metrics, improving accuracy, f_{mAP}, and F_1 to 90.16%, 95.24%, and 92.18%, respectively. Compared with the baseline ConvNeXt, PF-ConvNeXt increases these metrics by 4.74%, 5.46%, and 5.95%, respectively. In terms of complexity, PF-ConvNeXt employs 30.1 M parameters and 5.0 GFLOPs—higher than most lightweight models, yet comparable to or lower than several contemporary architectures while delivering substantially better recognition metrics. Inference speed measurements further validate its practicality. All models are benchmarked under identical hardware, input resolution, and a batch size of 1. Under these conditions, PF-ConvNeXt achieves 78 FPS, surpassing several recent models including QKFormer, GAC-SNN, and TLENet, which operate at 33 FPS, 28 FPS, and 24 FPS, respectively. Meanwhile, lightweight architectures such as MobileNet and ResNet50 offer higher frame rates of 96 FPS and 115 FPS, yet at the cost of considerably lower recognition accuracy. This trade-off underscores that PF-ConvNeXt not only enhances feature discrimination but also maintains competitive efficiency suitable for real-time deployment. The results consistently suggest that PF-ConvNeXt provides stronger discriminative capability for multi-class recognition under complex weather conditions, achieving higher predictive consistency and robustness with a reasonable computational overhead.

Figure 7 presents the confusion matrix results for various adverse-weather samples in the test set, which are used to further analyze the recognition performance of PF-ConvNeXt across different weather categories. Note that the confusion matrices are normalized for

visualization and are mainly used for qualitative error analysis; the overall quantitative comparison is reported in Table 5. Overall, the model achieves high classification accuracy for most categories, indicating that the proposed method provides strong general discriminative capability. Nevertheless, the recognition performance still varies across weather types.

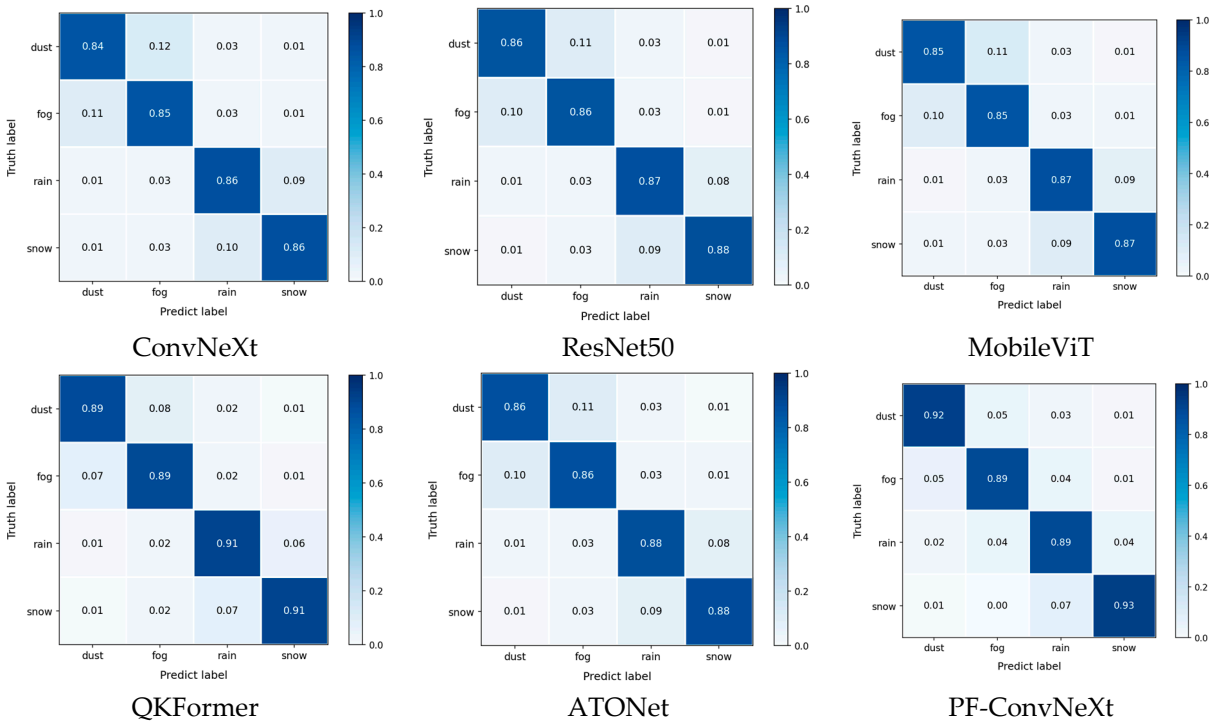

Figure 7. Confusion matrices of different models.

PF-ConvNeXt performs particularly well on weather conditions with strong global visual cues, such as dense fog or extensive sand/dust scenes. These conditions typically cause a global contrast reduction or clear shifts in color distribution. With PSA-based multi-scale feature fusion, the model captures such global degradation patterns effectively, facilitating reliable classification. In contrast, weather types characterized by subtle local cues and strong visual overlap with other categories are more difficult to distinguish and may be confused with normal weather or closely related classes. Confusion-matrix analysis suggests that these errors are largely driven by visual similarity across conditions, especially when weather effects appear as low-intensity local textures that overlap with underlying scene structures. In addition, intra-class variation, image sharpness differences, and augmentation choices may also affect recognition stability for some categories.

Overall, the confusion-matrix analysis confirms the effectiveness of PF-ConvNeXt for multi-class adverse-weather recognition. At the same time, it suggests that further improvements are possible for visually similar categories, which will be explored in future work.

5. Discussion

5.1. Limitations and Generalization Considerations

Although PF-ConvNeXt demonstrates strong classification performance under complex weather conditions, its generalization capability and potential overfitting risk warrant careful consideration. The dataset used in this study was expanded to 5000 images via augmentation; however, it originates from approximately 1500 raw images collected from three sources and covers only four adverse-weather categories (rain, snow, fog, and sandstorm). Compared with large-scale general-purpose datasets, the current dataset may not

fully capture the diversity of real-world driving scenes, including variations in weather intensity, illumination conditions, and geographical contexts. Such limitations in data scale and diversity may cause the model to over-learn dataset-specific patterns. To mitigate this risk, we incorporated several strategies in our experimental design. First, we employed data augmentation (random brightness adjustment, Gaussian noise, occlusion, and rotation) to increase input diversity and simulate practical perturbations. Second, we used regularization techniques, including weight decay and early stopping, to constrain model complexity and reduce over-optimization on the training set. Third, the proposed architectural components also contribute to robustness: the PSA module enhances tolerance to scale variations through multi-scale feature fusion, while the FECS module recalibrates channel and spatial responses to emphasize weather-related cues and suppress irrelevant background noise. The consistent improvements observed in the ablation studies further support the effectiveness of these designs in enhancing robustness. Nevertheless, extending the dataset with more diverse sources and conducting cross-domain evaluations remain important directions for future work.

5.2. Failure Cases and Error Analysis

The confusion matrices in Figure 7 demonstrate that while PF-ConvNeXt achieves reliable recognition for most weather categories, errors are concentrated in three scenarios: weak weather intensity, visual similarity between categories, and adverse illumination conditions. When rain or fog is extremely light, subtle cues are easily overwhelmed by background structures. For instance, a rain image with sparse streaks against a complex urban scene was misclassified as clear—the few rain streaks intertwined with building textures were difficult to isolate. Similarly, a thin fog image was mistaken for clear due to minimal contrast reduction. In such cases, the model relies more on dominant scene content than on weak weather features. Visual similarity between categories also leads to confusion, particularly between sandstorm and fog. A sandstorm image with a yellowish tint and mild degradation was misclassified as fog, as its global appearance resembled fog while fine-grained dust textures were not fully captured; the confusion matrix shows approximately 6% of sandstorm samples mislabeled as fog. Nighttime conditions further exacerbate misclassifications due to glare, reflections, and low-light noise. For example, a nighttime snow image was misclassified as rain, as snowflakes were obscured by bright streaks from vehicle lights, and a clear night scene with glare producing a veiling effect was mistaken for fog. These failure cases suggest that introducing fine-grained weather intensity annotations, decoupling global and local features, and incorporating temporal information or domain adaptation could further enhance model robustness.

5.3. Comparison with Related Studies and Practical Considerations

Compared with recent environment-condition recognition methods that emphasize lightweight deployment [15] or multi-label prediction [22], PF-ConvNeXt focuses on enhancing ConvNeXt with multi-scale fusion and spatial-channel attention to better capture both global degradation patterns and local textures. Relative to joint frameworks that couple weather recognition with downstream detection [26], our study targets a robust front-end weather classifier that can provide reliable priors for subsequent perception modules. Nevertheless, differences in datasets, label taxonomies, and evaluation protocols across studies may affect direct comparability. For practical deployment, efficiency-accuracy trade-offs should be considered: while PF-ConvNeXt provides strong recognition metrics, further optimization such as pruning, quantization, and knowledge distillation [34] could be explored to reduce latency and power consumption on embedded devices.

6. Conclusions

To meet the practical demand for adverse-weather recognition in real road and surveillance scenarios, this study proposes PF-ConvNeXt, an improved ConvNeXt-based model. The proposed approach enhances recognition performance under complex environments from three aspects: multi-scale feature modeling, feature enhancement, and training objective optimization. Specifically, we introduce a PSA multi-scale feature fusion module to jointly capture global and local weather cues. We further design FECS, a spatial–channel joint attention mechanism, to emphasize weather-related regions while suppressing background interference. In addition, Focal Loss is adopted to increase the learning emphasis on hard samples and minority classes, thereby alleviating training bias caused by class imbalance. Experiments are conducted on a dataset of 5000 images, constructed from RTTS, DAWN, and a self-collected rainy-weather dataset, and augmented proportionally to preserve the original class distribution. Under a unified experimental protocol, we perform both ablation and comparative evaluations. The ablation results demonstrate that PSA, FECS, and Focal Loss each contribute to performance improvements, and their combination achieves the best results. Comparative experiments further show that PF-ConvNeXt outperforms ConvNeXt, ResNet50, MobileNet, MobileViT, and FasterNet in terms of recognition accuracy and precision, validating the effectiveness and stability of the proposed architecture for multi-class adverse-weather recognition. Future work can be extended in several directions. First, broader real-world data covering more diverse scenes, weather severity levels, and illumination conditions should be collected to improve cross-region and cross-time generalization. Second, hard-sample-aware training strategies can be incorporated to further enhance discrimination of extreme weather and visually similar categories. Third, model lightweighting and inference acceleration should be explored while maintaining accuracy, to support real-time deployment in onboard systems.

Author Contributions: Q.W. contributed to software development, data curation, validation, conceptualization, and visualization. Z.Z. (Zhaofa Zhou) contributed to funding acquisition, formal analysis, and supervision. Z.Z. (Zhili Zhang) contributed to project administration, resources, and writing—review and editing. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Natural Science Foundation of China (Grant No. 62305393) and the Shaanxi Province Science and Technology Innovation Team Project (Grant No. 2025RSCXTD-046).

Data Availability Statement: The datasets and code generated and/or analyzed during this study are available from the corresponding author upon reasonable request.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. LeCun, Y.; Bottou, L.; Bengio, Y.; Haffner, P. Gradient-based learning applied to document recognition. *Proc. IEEE* **1998**, *86*, 2278–2324. [CrossRef]
2. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet classification with deep convolutional neural networks. *Commun. ACM* **2017**, *60*, 84–90. [CrossRef]
3. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv* **2014**, arXiv:1409.1556.
4. Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; Rabinovich, A. Going deeper with convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 1–9.
5. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhin, I. Attention is all you need. *arXiv* **2017**, arXiv:1706.03762.
6. Liu, Z.; Mao, H.; Wu, C.Y.; Feichtenhofer, C.; Darrell, T.; Xie, S. A convnet for the 2020s. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 19–24 June 2022; pp. 11976–11986.

7. Zhang, Y.; Carballo, A.; Yang, H.; Takeda, K. Perception and sensing for autonomous vehicles under adverse weather conditions: A survey. *ISPRS J. Photogramm. Remote Sens.* **2023**, *196*, 146–177. [CrossRef]
8. Gupta, H.; Kotlyar, O.; Andreasson, H.; Lilienthal, A.J. Video weather recognition (VARG): An intensity-labeled video weather recognition dataset. *J. Imaging* **2024**, *10*, 281. [CrossRef] [PubMed]
9. Karvat, M.; Givigi, S. Adver-city: Open-source multi-modal dataset for collaborative perception under adverse weather conditions. *arXiv* **2024**, arXiv:2410.06380.
10. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
11. Huang, G.; Liu, Z.; Van Der Maaten, L.; Weinberger, K.Q. Densely connected convolutional networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 4700–4708.
12. Shigesawa, A.; Yagi, M.; Takahashi, S.; Takedomi, S.; Mori, T. Winter road surface condition recognition in snowy regions based on image-to-image translation. *Sensors* **2025**, *26*, 241. [CrossRef] [PubMed]
13. Howard, A.G.; Zhu, M.; Chen, B.; Kalenichenko, D.; Wang, W.; Weyand, T.; Andreetto, M.; Adam, H. MobileNets: Efficient convolutional neural networks for mobile vision applications. *arXiv* **2017**, arXiv:1704.04861. [CrossRef]
14. Chung, Y.L. Efficient lightweight image classification via coordinate attention and channel pruning for resource-constrained systems. *Future Internet* **2025**, *17*, 489. [CrossRef]
15. Manivannan, P.; Sathyaprakash, P.; Jayakumar, V.; Chandrasekaran, J.; Ananthanarayanan, B.; Sayeed, M. Weather classification for autonomous vehicles under adverse conditions using multi-level knowledge distillation. *Comput. Mater. Contin.* **2024**, *81*, 4327. [CrossRef]
16. Introvigne, M.; Ramazzina, A.; Walz, S.; Scheuble, D.; Bijelic, M. Real-time environment condition classification for autonomous vehicles. In Proceedings of the 2024 IEEE Intelligent Vehicles Symposium, Sorrento, Italy, 2–6 June 2024; pp. 1527–1533.
17. Dosovitskiy, A. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv* **2020**, arXiv:2010.11929.
18. Touvron, H.; Cord, M.; Douze, M.; Massa, F.; Sablayrolles, A.; Jégou, H. Training data-efficient image transformers & distillation through attention. In Proceedings of the International Conference on Machine Learning, Virtual, 18–24 July 2021; pp. 10347–10357.
19. Liu, Z.; Lin, Y.; Cao, Y.; Hu, H.; Wei, Y.; Zhang, Z.; Lin, S.; Guo, B. Swin transformer: Hierarchical vision transformer using shifted windows. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Virtual, 11–17 October 2021; pp. 10012–10022.
20. Cui, J.; Chen, Y.; Wu, Z.; Wu, H.; Wu, W. A driver behavior detection model for human-machine co-driving systems based on an improved Swin Transformer. *World Electr. Veh. J.* **2024**, *16*, 7. [CrossRef]
21. Wang, W.; Xie, E.; Li, X.; Fan, D.P.; Song, K.; Liang, D.; Lu, T.; Luo, P.; Shao, L. Pyramid vision transformer: A versatile backbone for dense prediction without convolutions. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Virtual, 11–17 October 2021; pp. 568–578.
22. Chen, S.; Shu, T.; Zhao, H.; Tang, Y.Y. MASK-CNN-Transformer for real-time multi-label weather recognition. *Knowl.-Based Syst.* **2023**, *278*, 110881. [CrossRef]
23. Wu, H.; Xiao, B.; Codella, N.; Liu, M.; Dai, X.; Yuan, L.; Zhang, L. CvT: Introducing convolutions to vision transformers. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Virtual, 11–17 October 2021; pp. 22–31.
24. Dai, Z.; Liu, H.; Le, Q.V.; Tan, M. CoatNet: Marrying convolution and attention for all data sizes. In Proceedings of the Advances in Neural Information Processing Systems, Virtual, 6–14 December 2021; pp. 3965–3977.
25. Mehta, S.; Rastegari, M. MobileViT: Light-weight, general-purpose, and mobile-friendly vision transformer. *arXiv* **2021**, arXiv:2110.02178.
26. Aloufi, N.; Alnori, A.; Basuhail, A. Enhancing autonomous vehicle perception in adverse weather: A multi objectives model for integrated weather classification and object detection. *Electronics* **2024**, *13*, 3063. [CrossRef]
27. Marathe, A.; Ramanan, D.; Walambe, R.; Kotecha, K. Wedge: A multi-weather autonomous driving dataset built from generative vision-language models. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Vancouver, BC, Canada, 18–22 June 2023; pp. 3318–3327.
28. Zhang, H.; Zu, K.; Lu, J.; Zou, Y.; Meng, D. EPSANet: An efficient pyramid squeeze attention block on convolutional neural network. In Proceedings of the Asian Conference on Computer Vision, Macau, China, 4–8 December 2022; pp. 1161–1177.
29. Lin, T.Y.; Goyal, P.; Girshick, R.; He, K.; Dollár, P. Focal Loss for dense object detection. In Proceedings of the 2017 IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 2980–2988.
30. Li, B.; Ren, W.; Fu, D. Benchmarking single-image dehazing and beyond. *IEEE Trans. Image Process.* **2018**, *28*, 492–505. [CrossRef] [PubMed]
31. Kenk, M.A.; Hassaballah, M. DAWN: Vehicle detection in adverse weather nature dataset. *arXiv* **2020**, arXiv:2008.05402. [CrossRef]
32. Chen, J.; Kao, S.H.; He, H.; Zhuo, W.; Wen, S.; Lee, C.H.; Chan, S.H.G. Run, don't walk: Chasing higher FLOPS for faster neural networks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Vancouver, BC, Canada, 18–22 June 2023; pp. 12021–12031.

33. Zhou, C.; Zhang, H.; Zhou, Z.; Yu, L.; Huang, L.; Fan, X.; Tian, Y. QKFormer: Hierarchical spiking transformer using QK attention. In Proceedings of the Advances in Neural Information Processing Systems, Vancouver, BC, Canada, 9–15 December 2024; pp. 13074–13098.
34. Shin, H.; Choi, D.W. Teacher as a lenient expert: Teacher-agnostic data-free knowledge distillation. In Proceedings of the AAAI Conference on Artificial Intelligence, Vancouver, BC, Canada, 20–27 February 2024; pp. 14991–14999.
35. Qiu, X.; Zhu, R.J.; Chou, Y.; Wang, Z.; Deng, L.J.; Li, G. Gated attention coding for training high-performance and efficient spiking neural networks. In Proceedings of the AAAI Conference on Artificial Intelligence, Vancouver, BC, Canada, 20–27 February 2024; pp. 601–610.
36. Wu, X.; Gao, S.; Zhang, Z.; Li, Z.; Bao, R.; Zhang, Y.; Wang, X.; Huang, H. Auto-Train-Once: Controller network guided automatic network pruning from scratch. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 17–21 June 2024; pp. 16163–16173.