ITC 3/54
Information Technology and Control
Vol. 54 / No. 3/ 2025
pp. 885-899
DOI 10.5755/j01.itc.54.3.41307

REWeather: A Unified Detection Framework for Automatic Driving Images Restoration and Enhancement in Adverse Weather

Received 2025/04/24 | Accepted after revision 2025/06/10

HOW TO CITE: Zhang, X., Zhang, X., Peng, X., Dong, M., Ota, K., Zhang, X. (2025). REWeather: A Unified Detection Framework for Automatic Driving Images Restoration and Enhancement in Adverse Weather. *Information Technology and Control,* 54(3), 885-xxx. https://doi.org/10.5755/j01.itc.54.3.41307

REWeather: A Unified Detection Framework for Automatic Driving Images Restoration and Enhancement in Adverse Weather

Xiaoyu Zhang*
School of Artificial Intelligence, Shenyang University of Technology, 111 Shenliao West Road, Shenyang, 110870, Liaoning, China; e-mail: xy.zhang@sut.edu.cn
Shenyang Key Laboratory of Industrial Intelligent Chip and Network System Innovation Application.

Xinyu Zhang, Xiting Peng*
School of Information Science and Engineering, Shenyang University of Technology, 111 Shenliao West Road, Shenyang, 110870, Liaoning, China; e-mail: fhks1999@smail.sut.edu.cn; xt.peng@sut.edu.cn

Mianxiong Dong, Kaoru Ota
Department of Sciences and Informatics, Muroran Institute of Technology, 27-1 Mizumoto-cho, Muroran, 050-8585, Hokkaido, Japan; e-mail: mx.dong@csse.muroran-it.ac.jp; ota@csse.muroran-it.ac.jp

Xiaoling Zhang
School of Information Science and Engineering, Shenyang University of Technology, 111 Shenliao West Road, Shenyang, 110870, Liaoning, China; e-mail: zhangxiaoling@sut.edu.cn
Shenyang Key Laboratory of Industrial Intelligent Chip and Network System Innovation Application.

Corresponding author: xy.zhang@sut.edu.cn; xt.peng@sut.edu.cn

Recently, with the rapid development of autonomous driving technology, it prompts the vehicle detection technology to continuously improve its accuracy, stability and reliability to better meet the needs of self-driving. However, due to the interference of adverse factors in adverse weather, the decrease of detection accuracy of autonomous vehicle is led to the phenomenon of missing and wrong detection, which has a serious impact on the safety of autonomous vehicles. Therefore, we propose REWeather to solve such problems of autonomous vehicles in multiple adverse weather conditions. Firstly, to classify the types of adverse weather, distinguishing

among foggy, rainy and snowy weather, Broad Learning System (BLS) which is simple and efficient is used in REWeather. Due to the impact of these adverse weathers on sensors, simple dark channel and guided filtering methods is used to preprocess foggy and rainy images, respectively. Then, we put the processed images into the Real-Enhanced Super-Resolution Generative Adversarial Networks (Real-ESRGAN) for further denoising and enhancing the details of detected objects, enabling the sensor to recognize other targets on the road faster and better in adverse weather. To ensure the best detection results, we also use latest Realtime Detection Transformer (RT-DETR) as the detector to validate our work and the final model is deployed on the edge device. Moreover, we use several public datasets and our own collected data to make a real world dataset containing a variety of adverse weathers to train and test our proposed framework, which makes it closer to the real situation. The results demonstrate that our framework achieves a 3.8% improvement in mAP, significantly enhancing the detection capability of autonomous vehicles under adverse weather conditions.

KEYWORDS: Vehicles Detection, Adverse Weather Removal, Broad Learning System, Realtime Detection Transformer, Super-Resolution Generative Adversarial Networks

1. Introduction

With the progressive progress of artificial intelligence technology, autonomous vehicle technology is becoming mature gradually. Effective detection of objects on the lane is a fundamental prerequisite for the implementation of self-driving. Benefiting from advancement of object detection technology, current mainstream detectors for autonomous driving have shown good results in some simple daily situations [30]. However, as the application scenarios of autonomous vehicles become more and more complex, these detectors often fail to achieve the ideal detection effect in the face of complex scenarios or without large enough data for training. Moreover, existing mainstream detecting methods are biased towards clear weather conditions mostly. Due to various adverse factors on the sensor image interference, especially in bad weather such as heavy snow, dense fog and rain, which lead to blurred details on the target in the lane that are difficult to detect [25]. Therefore, the adverse effects of adverse weather conditions on autonomous driving images have led to missing and false detection, resulting in safety hazards in autonomous driving. Thus, it is especially important to research sensor performance in inclement weather for the safety of autonomous vehicle technology.

Firstly, the restoration and enhancement of corresponding images are crucial for improving the safety of autonomous driving in adverse weather conditions. Rainy, foggy, and snowy weather can result in the functionality of cameras and lidar to decrease significantly [4]. Most of existing 2D detection methods are aimed at single weather or single situation under adverse weather and propose the corresponding denoising and restoring the deteriorated image solution. For snow conditions, recent approaches have employed snow particle segmentation and multi-spectral fusion techniques to address the unique challenges posed by snow accumulation and reflection. Hodges et al. [8] proposed a single image dehazing method based on deep neural network. A Rain Intensity Controlling Network (RIC-Net) is presented by Ni et al. [14]. It consists of three sub-networks: a background extraction network, a high-frequency rain-streak elimination network, and the main controlling network. This allows for continuous interpolation control of rain images with varying intensities. To eliminate rain streaks from a picture, Fu et al. [5] used a deep convolutional neural network (CNN). However, most of these methods only provide solutions for a specific type of severe weather, and are lacking in universality in the face of realistic and changeable weather conditions.

On the other hand, in addition to improving and enhancing the image, the accuracy and real-time performance of the detector are also crucial for maintaining the safety of autonomous driving in adverse weather conditions. SINet [9] has got fast detection via implementing a scale insensitive network. However, most detection models cannot meet the high requirements for both accuracy and real-time speed in specific adverse weather conditions. In addition,

database under adverse weather conditions is difficult to obtain, many methods use synthetic data to train and test model. Sakaridis et al. [19] used a model based on CNN to add synthetic fog on real road traffic images to study the de-fog algorithm in traffic environment. Although these existing methods for synthesizing adverse weather data can fill the amount of data to a certain extent, the trained model is still different from the real degradation of images in real adverse weather. Relatively, models trained using real data are more stable and robust. Therefore, a framework that can solve the problem of false detection in automatic driving detection under real-world various bad weather conditions is really important to enhance the safety of automatic driving.

With above motivations, we propose the framework which can reduce noise dete-rioration of images under adverse weather and enhance details of vehicles features to improve detection accuracy, and use the existing public data set and the data we collected ourselves to make a non-synthetic data set containing three kinds of real adverse weather scenarios (details in Section 4). REWeather framework use the lightweight network BLS proposed by Chen et al. [2] to classify the real adverse weather images quickly, and then carry out different processing according to different adverse weather types, then put them in Real-ESRGAN proposed by Wang et al. [22] to carry out super-resolution enhancement processing, and the newest RT-DETR is used as final unified detection. BLS model classification model has the ability to quickly and accurately classify different kinds of adverse weather. After classification, we can do different processing on the deterioration of different kinds of weather. All data are classified by BLS into three categories: rainy, snowy and foggy. Then we use dark channel fog removal [7] to process fog images to lessen the impact of fog on the road target, and use guided filtering [24] to process rainy images to reduce the influence of raindrops. This is because although Real-ESRGAN has good capability of image restoration and feature enhancement, if it directly enhances the image of bad weather, the interference of unfavorable factors in the original image will be amplified, so it is necessary to classify and carry out different simple processing. Taking into account the real-time requirements of automatic driving, RT-DETR is used as the detector of the REWeather. Finally, we deploy the REWeather framework on edge device, ensuring the practicality of our framework. The goal of the paper is to combine unified data processing and detection for different kinds of real adverse weather, as well as deployment on edge-end devices. The main contributions of this paper are as follows:

- We combine Real-ESRGAN and RT-DETR for the first time and use the frame- work for automatic driving detection under adverse weather, effectively reducing the deterioration of detection images caused by various adverse factors in bad weather and improving the problem of wrong detection and missing detection of autonomous vehicles.
- For the first time, we use RT-DETR for autonomous driving in adverse weather. The lightweight network Broad Learning System is used to classify three kinds of adverse weather, and different processing methods are carried out according to different adverse weather. The REWeather can be deployed on the edge device, which reduces the occupation of resources and improves the efficiency of automatic driving.
- Experiments show that, based on enhanced real and non-synthetic data under various adverse weather conditions, our framework improves the detection ability of 2D sensors of autonomous vehicles under adverse weather, which is 3.8% higher than that of mAP using unprocessed training data. In addition, the BLS classification network has more advantages in training time than other deep learning networks. Moreover, RT-DETR has higher accuracy than conventional mainstream detectors.

To better describe our framework: Section 2 is used to summarize the related work of this paper. The detailed model will be presented in Section 3. Section 4 is comprehensive and details experiments about the framework in this paper. Finally, a conclusion is concluded by us in Section 5.

2. Related Work

2.1. Broad Learning System

Due to the large amount of computation and high computational cost of deep learning, BLS, a lightweight network structure is proposed to solve this problem. Compared with traditional deep network structure, BLS does not need complex network

depth, and enhance the network ability by increasing the network width, that is, increasing feature nodes, rather than deep learning continuously increasing the depth of the network. Secondly, when the training data increase, deep learning network usually needs to retrain the whole network model, while BLS can use incremental learning. The system can be transformed by incremental learning without retraining the model, which greatly saves training time and improves the efficiency of network learning. BLS is widely used in real-world classification applications due to its efficient and concise network structure. BLS is utilized by Zhao et al. [26] to achieve fault detection, which effectively completed the rotor system fault classification. Peng et al. [15] proposed a lightweight vehicle classification method that used BLS. The results show that the training speed of this method can be increased by almost 10 times compared to general CNN- based methods. In summary, BLS has great potential in scenarios that require fast and accurate classification.

2.2. Adverse Weather Removal and Enhancement

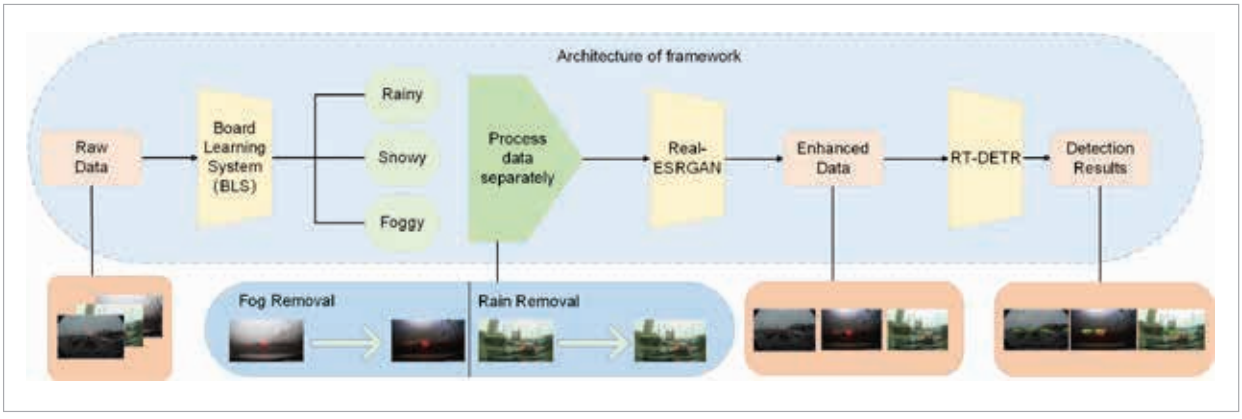
2.2.1. Single Adverse Weather Image Removal

Adverse weather such as dense fog, heavy snow and torrential rain reduce visibility and destroy the information captured by the images. This greatly affects the performance of computer vision systems. Therefore, to make these computer vision systems more reliable, adverse weather effects must be removed from the images. Early methods of eliminating adverse weather effect included prior modeling of weather conditions using empirical observations. These priors must be modeled separately for each weather condition and a common prior for all invalid weather conditions. For example, He et al. [7] suggested a technique for dark channel dehazing that is both straightforward and efficient. Xu et al. use a guidance filtering method [24] using special guidance images to remove noise from rainy images. In addition to this, recently CNN-based methods have also been widely researched for removing snowfall, fog, and raindrops. However, the simple processing methods are not sufficient to cope with complex severe weather degradation situations.

2.2.2. Super-resolution Generative Adversarial Network

Super-resolution Generative Adversarial Network (SRGAN) [11] makes GAN have the ability of super-resolution firstly. Later studies based on the model improvement of SRGAN proposed ESRGAN [21] and Real-ESRGAN. Real-ESRGAN is the latest super-resolution GAN with strong super-resolution ability proposed in recent years. To replicate the real deterioration, Real-ESRGAN suggested a higher-order degradation mechanism. Furthermore, Real-ESRGAN can recover actual photos, which is more applicable in the real world than previous super-resolution models. Super-resolution networks have gained popularity recently for practical uses. Zheng et al. [27] used Real-ESRGAN to enhance cropped forest fire images to secure timeliness and accuracy of forest fire monitoring. Zhu et al. [29] enhanced the super-resolution image effect and reduced training

Figure 1
Architecture of the proposed framework.



time by changing the degradation process and discriminator. However, this enhancement method is not suitable for severely degraded original images, as directly processing the original image would actually increase the impact of adverse factors on the image.

2.3. Object Detection in Automatic Driving

Object detection technology is the core content of automatic driving. How to detect other vehicles quickly and accurately is the key to realize the safety and reliability of automatic driving technology. Since deep learning has advanced, the progress of object detection technology also makes automatic driving technology more mature.

Nowadays, object detection algorithms such as faster region convolutional neural network (Faster R-CNN) [18], You-Only-Look-Once (YOLO) [17] and Single shot multibox detector (SSD) [13] are widely used in the field of automatic driving. Wang et al. [23] proposed a universal object detection framework called UniDetector which can augment the generalization ability of framework and solve problem of detection in real-world complex scenarios. In addition, end-to-end object detection with transformers [1] also have gained great attention recently due to their simpler model framework and excellent detection performance. CNN is used to extract features immediately for classification and regression in one-stage algorithms like YOLO and SSD, which equally carry out extensive sampling at diverse parts of the picture and adopt varying sizes and aspect ratios, so its speed is fast, but the accuracy is relatively low. On the other hand, the two-stage method, such as Faster-RCNN, first generates sparse candidate boxes, and then classifies and does the regression of them, which has higher accuracy but lacks of real-time detection capability. DETR predicts the final detection result directly (in parallel) by combining the common CNN with the transformer architecture [20]. Deformable DETR [28] is intended to address these problems by focusing its attention module solely on a limited number of significant sampling points surrounding the reference. With ten times less training time than DETR, deformable DETR can outperform DETR, particularly on tiny targets. Group DETR [3] introduces multiple object queries to retain the advantages of DETR's end-to-end reasoning while taking advantage of the one-to-many advantages in training to improve performance and speed up model convergence. Considering the background of autonomous driving, in addition to the requirement for accuracy, the real-time performance of the detector is also crucial.

The above methods are only for a specific link in the actual situation, lacking effective combination in real-world application. Different from the above methods, our REWeather takes advantage of the advantages of various methods and makes up for their respective shortcomings by combining them. The framework realizes the effective detection of road targets by autonomous vehicles in a variety of adverse weather scenarios.

3. System Framework

In this section, the details of REWeather framework are given in Figure 1. BLS in classification module is used to classify different kinds of adverse weather images so that it can perform different de-noising processes before uniformly putting into Real-ESRGAN features enhancements in adverse weather removal block. Simple network structure, fast and effective classification capability and incremental learning of BLS are easy to expand meet our needs for fast classification of adverse weather, so BLS is used as the classifier of our framework. Because different kinds of adverse weather will cause different kinds of effects on the automatic driving images obtained by the sensor, if the same super-resolution enhancement is carried out on these images directly, the impact of adverse factors on the images will be increased, so we need to classify different kinds of adverse weather. Next, simple processing is carried out for different kinds of bad weather images. Dark channel is used to de-fog images in foggy days, and guided filtering to reduce the interference of raindrops in rainy days. In addition, the test proved that the snowy images without preliminary processing and directly take super-resolution enhancement still has good results, so we did not do another processing on the snowy images. After the above steps, Real-ESRGAN is used to enhance the features and further de-noising the adverse weather images. Due to its powerful super-resolution capability, the features of lane targets, such as pedestrians and vehicles are greatly enhanced and easier to detect. Finally, RT-DETR is used as the detector of REWeather.

3.1. Classification Module

In daily life, autonomous vehicles will encounter a variety of adverse weather factors on the road, and different weather interferes with sensors in different ways. Therefore, it is very important to quickly classify adverse weather and handle corresponding different situations. Nowadays, CNN-based methods have shown good results, but due to the characteristics of deep learning, the training is time-consuming, and when the network requirements change, it has to be retrained, which is not conducive to expansion. BLS is a flat network consisting of mapped features and enhancement nodes, which is easier to extend. We can see BLS network from Figure 2. For the model in this paper, the images of adverse weather are transformed into the matrix after a series of data processing, which is denoting O as the input data, and the pseudo-inverse of the feature nodes and enhancement nodes to the goal value is then calculated to provide output data P. Input data O is a matrix of n × m, n represents the number of samples of input, and m represents the dimension of each sample.

Firstly, the input O is processed for forming the mapping node *Jz*. Each mapping node can generate q feature mappings, and each feature mapping has w corresponding nodes. At first, input O is mapped into J_z by Equation (1):

$$J_z = \lambda_z\left(OI_{\alpha_z} + \chi_{\alpha_z}\right), \quad z = 1,2,3,\ldots,q, \tag{1}$$

where $I_{\alpha z}$ and $\chi_{\alpha z}$ are both generated randomly. λ_z is a random mapping function which can be determined by input data or feature mapping. Then, from Equation (1), we can get $J^q = [J_1,J_2,J_3,\ldots,J_t]$, which contains all mapping features extracted from the input data O.

Subsequently, based on J^q and the I_{ex} and χ_{ex} generated randomly, next enhancement nodes use nonlinear functions μ_x in the following Equation (2):

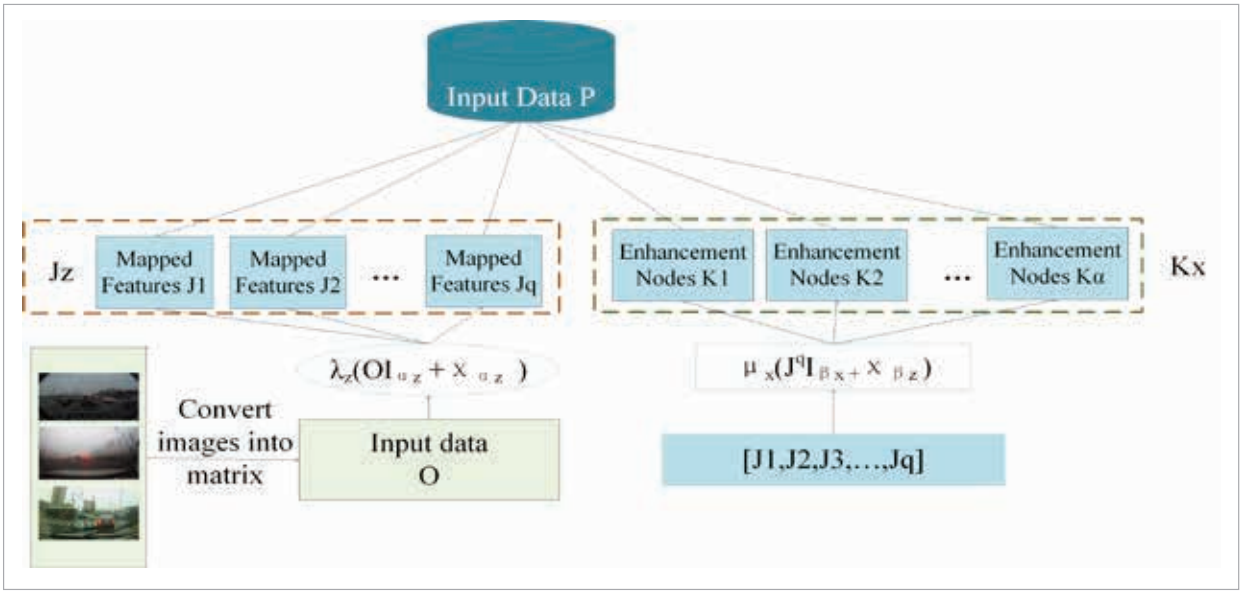
$$K_x = \mu_x\left(J^q I_{\beta_x} + \chi_{\beta_x}\right), \quad x = 1,2,3,\ldots,\alpha, \tag{2}$$

where each K_x calculated by Equation (2) is combined to $K^\alpha = [K_1,K_2,K_3,\ldots,K_\alpha]$, rep- esenting the enhancement nodes extracted from the previously obtained mapping features.

At last, the mapped J^q features and extracted K^α enhancement nodes are combined into a matrix, and then this matrix is multiplied with the network weight I to get the final output data *P*, as in the following Equation (3):

$$P = [J^q|K^\alpha]I = UI, \tag{3}$$

Figure 2
Architecture of Broad Learning System.



where U is matrix combined by the mapped J^q features and extracted K^α enhancement nodes, and I as the weight, which can be expressed as follows in Equations (4)-(5):

$$(U)^+ = \lim_{n\to 0}\left(n + U^Q U\right)^{-1} U^Q \tag{4}$$

$$I = (U)^+ Z\,. \tag{5}$$

Normally, n 0 is needed, V is unit matrix, $(U)^+$ is used to calculate the pseudo inverse of matrix U. Next, we obtain the weight I by multiplying it by matrix Z. From the above steps, we can obtain the final desired output data P.

Except for this, after we have trained the basic model, incremental learning can be used for extending the network which need not retrain the whole model that saves resource usage and enhances the augmentability of network. It can be expressed as follows in Equation (6):

$$P^{\alpha+1} = \left[P\middle|\mu\left(J^q I_{\beta_{x+1}} + \chi_{\beta_{x+1}}\right)\right], \tag{6}$$

Algorithm 1: System framework

Input: Adverse Weather Images I_{ad} contains rainy images I_r, foggy images I_f and snowy images I_s

Output: Enhanced images after detection $D(I_{Ead})$

Set $I_{ad} = \{i_{ad_n} | n = 1, 2, 3, ..., n\}$.
for $i_{ad_n} \leftarrow 0$ to I_{ad} **do**
if i_{ad_n} in I_r **then**
    Noise removal by guided images filtering method;
    Enhancement by Real-ESRGAN;
    return I_{Er};
else if i_{ad_n} in I_f **then**
    Noise removal by dark channel method;
    Enhancement by Real-ESRGAN;
    return I_{Ef};
else
    Enhancement by Real-ESRGAN;
    return I_{Es};
end if
end for
Set $I_{Ead} = \{I_{Er}, I_{Ef}, I_{Es}\}$.
Set $i_{Ead} = \{i_{Ead_m} | m = 1, 2, 3, ..., n\}$.
for i_{Ead_m} 0 to I_{Ead} **do** 19: Detected by RT-DETR;
    return D(I_{Ead});
end for

where β_{x+1} and $\chi_{\beta_{x+1}}$ are also randomly generated matrices. By calculating pseudo- inverses in the same way, we can get a new model after using incremental learning without retraining the whole network.

3.2. Adverse Weather Removal and Enhancement Module

In this section, we will provide a detailed introduction to dark channel dehazing, guided filtering for rain removal, and the structural principles of Real-ESRGAN. We will discuss how to restore and enhance features of severe weather images, remove real-world noise, and enhance the detailed features of road detection objects. Due to the particularity of foggy and rainy images, direct super-resolution enhancement will increase the interference of noise. This is because in foggy and rainy weather, fog occlusion and raindrop rain marks have a serious impact on road targets. Although Real-ESRGAN has strong enhancement ability, directly enhancing unprocessed image information in severe weather images will actually increase the influence of adverse factors on the image. Therefore, simple and effective dark channel dehazing algorithm and guided filtering algorithm are used to reduce the impact of thick fog and rainfall on the image to a certain extent, and then uses Real-ESRGAN to uniformly enhance the preliminary processed image and snowy image together.

Certain pixels will always have at least one low-value color channel in the majority of non-sky picture zones. The mathematical definition of this concept is as follows, in Equation (7):

$$R^{\text{dark}}(i) = \min_{c\in\{r,g,b\}}\left(\min_{g\in\Omega(i)}\left(R^c(J)\right)\right), \tag{7}$$

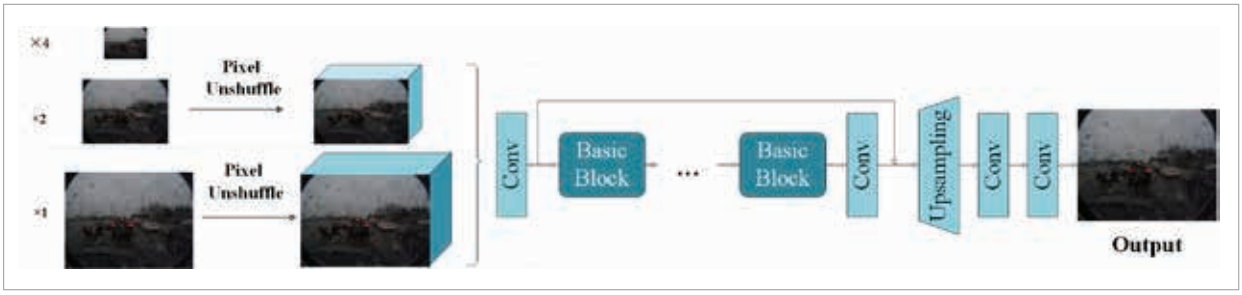
where R^c represents each channel of color image, $\Omega(i)$ is a local patch centered at i.

A flalign commonly used to describe images with fog is expressed as follows in Equation (8):

$$F(i) = R(i)s(i) + E\left(1 - s(i)\right), \tag{8}$$

where $F(i)$ represents fog image, $R(i)J$ represents fog-free image to be restored, E is the global atmospheric light component, and $s(i)$ is the transmittance.

Figure 3
Real-ESRGAN generator network architecture.



According to Equations (7)-(8), the final Equation (9) is as follows:

$$R(i) = \frac{F(i) - E}{\max(s(i), s_0)} + E\,, \tag{9}$$

where t_0 represents the lower bound of transmittance.

Through the above steps, we can achieve a simple fog removal result. Similarly, we use guided image filtering to reduce the impact of noise on images in rainy days. Guided filtering has very low time complexity and can effectively smooth the image. The method of guided filtering is using a guided image to generate weight, so as to process input image, and it can be expressed as follows in Equation (10):

$$l_g = \sum_h W_{gh}(G_i) \cdot R_h\,, \tag{10}$$

where l, G_i, R represents output image, guide image, and input image; g, h represents index of pixels in the image. It can be seen that the weight W in the above Equation (10) is only related to the guide image G_i, while the weight W in the bilateral filtering is determined by input image itself.

The central assumption of the guide filter is that output and the guide image are locally linear models in the local window w_k, the linear relationship between them can be expressed as follows in Equation (11):

$$l_g = a_k G_{ig} + b_k, \forall g \in w_k\,, \tag{11}$$

where a_k, b_k represents constant coefficient to be obtained by calculation. The formula is very simple but has the ability to retain the edge and the ability to smooth the noise, so we use this method in our framework.

Next, we put the after fog and rain removal images and snowy images into Real- ESRGAN for unified enhancement which is used to enhance detailed features of vehicles. Generator of Real-ESRGAN builds the higher-order degradation process of the image, which can be better used in real-world application scenarios that shown in Figure 3. The generator uses pixel shuffle to reduce spatial size and expand channel size, and then feeds the input into the network. Therefore, most calculations are performed in a smaller resolution space, effectively reducing the consumption of GPU memory and computing resources. The traditional degradation model cannot adapt well to the complexity and variability of real-world degradation, such as inevitable compression noise caused by image transmission over the network. In order to solve these problems, Real-ESRGAN proposed the high-order degradation model as follows, in Equation (12):

$$B(G) = \left[(G \otimes V) \downarrow_r + N\right]_{jpeg}\,, \tag{12}$$

where G is unprocessed adverse weather image, V is the blur function, $\downarrow_r$ is the factor about downsampling, N is the noise and $[]_{jpeg}$ is the adverse weather image compressed using the JPEG method. Equation (13) below gives a higher-order degradation process similar to Equation (12), Equation (13) is as follows:

$$L = B^S(G) = (B_S \ldots B_1)(G)\,, \tag{13}$$

where L represents output of the degradation model and S represents the number of degradation steps.

In addition, to solve the problem of ringing artifacts

near the sharp edges in images, Real-ESRGAN uses a sinc filter, and the filter kernel can be expressed as follows in Equation (14):

$$k(m,n)=\frac{s_c}{2\pi\sqrt{\left(m^2+n^2\right)}}K_1\left(s_c\sqrt{\left(m^2+n^2\right)}\right), \quad (14)$$

where (m,n) denotes the kernel coordinate of the filter, s_c denotes truncation frequency, and K_1 denotes first-order Bessel function.

3.3. Detection Module

Finally, we put the processed data into the RT-DETR for training and detecting which has higher accuracy and faster reaction time.

DETR employs a CNN-based backbone for feature extraction and dimension reduction. The output feature maps are reshaped, with each token's high dimensionality first reduced via 1×1 convolution before entering the transformer. The self-attention mechanism performs global analysis on these features, where larger objects benefit from enhanced spatial relationship modeling. Positional encodings, added before each multi-head self-attention layer, incorporate 2D spatial information through separately computed horizontal/vertical components.

The decoder uses a fixed set of learnable tokens for parallel prediction. Unlike traditional methods generating redundant proposals, DETR directly outputs a fixed number of predictions with each decoded token processed by shared FFNs to predict normalized box coordinates and class scores. RT-DETR enhances this architecture by optimizing encoder-decoder efficiency, achieving real-time performance while maintaining accuracy.

First, RT-DETR extracts three scale outputs from the backbone network, whose output steps are 8, 16 and 32, and we label them with M_3, M_4 and M_5. RT-DETR uses a layer of transformer encoder to process only the features outputed by backbone network. It turns the two-dimensional M_5 features into vectors and then hands them to the AIFI module in Equation (15) for processing. Then, we change the output back to two dimensions N_5 features, so as to complete the subsequent cross-scale feature fusion.

$$\begin{aligned} Q=K=V&=Flatten(M_5)\\ N_5&=Reshape\left(Attn(Q,K,V)\right)\end{aligned}, \quad (15)$$

where *Attn* represents the multi-head self-attention, and *Reshape* represents restoring the shape of the feature to the same as M_5, which is the inverse operation of Flatten.

The first target query for the decoder is then chosen by RT-DETR using IoU-Aware query selection based on a predetermined set of picture characteristics from the encoder output sequence. Constricting the model to provide a high classification score for features with high IoU scores and a low classification score for features with low IoU scores during training is how IoU-aware query selection is accomplished. As a result, both the classification score and the IoU score of the prediction box that corresponds to the Top-K encoder features based on the classification score are high. The optimization objective of the detector is expressed in Equation (16):

$$\begin{aligned} L(\hat{r},r)&=L_{box}(\hat{w},w)+L_{cls}(\hat{e},\hat{w},r,w)\\ &=L_{box}(\hat{w},w)+L_{cls}(\hat{e},e,IOU)\end{aligned}, \quad (16)$$

where r_{hat} and r represent forecast and ground truth, $r_{hat}=\hat{e},\hat{w}$ and $r=c,b,c$ and b represent categories and bounding boxes, respectively.

Figure 4
Edge facility Huawei Atlas 200I DK A2.

To achieve consistency restrictions on the classification and localization of positive samples, IoU scores are incorporated into the classification branch's objective function.

Finally, a decoder with an auxiliary prediction head iteratively optimizes the object query to generate boxes and confidence scores.

4. Experiment

In this section, we introduce the self-driving dataset under adverse weather used in our paper, evaluate the BLS in our framework by contrast experiments, and verify the effectiveness of our framework by ablation experiments. BLS experiments are based on the server with I5-12500H CPU and RTX 3060 laptop GPU. Detection experiments are based on the cloud server with NVIDIA RTX A4000 and Intel CPU E5-2686 v4. The test part of the experiment is carried out on the Huawei edge device Atlas 200I DK A2 which features an AI computing capability of 8 TOPS (INT8) and 4 TFLOPS (FP16), complemented by a 4GB LPDDR4X memory module. The device is equipped with a quad-core CPU operating at 1.0 GHz, and supports multiple interfaces including HDMI and USB 3.0. With a compact form factor of 44mm × 135mm × 120mm and a typical power consumption of 40W, this hardware platform demonstrates an optimal balance between computational performance and energy efficiency for embedded autonomous driving applications, as shown in Figure 4.

HUAWEI Atlas 200I DK A2 is based on the Ascend processor and is used for developing and deploying AI applications. Deploying the framework of this chapter on edge devices to simulate situations where actual computing resources are limited, in order to verify the application potential of the framework of this chapter.

4.1. Actual Adverse Weather Database

REWeather is aimed at the detection of different kinds of adverse weather in real-life situation, but most of the existing datasets of adverse weather autonomous driving are for one weather or a single condition, and many datasets use synthetic data, which cannot meet the complex situation of most of the real adverse weather.

Therefore, from the existing single adverse weather datasets and our own collected foggy images, we make a real-life adverse weather dataset containing foggy, rainy, snowy days. Among them, the data of snow days is from the images of snow days in the DAWN dataset [10] and part of the images of snow or containing large interference of snow conditions from the CADC dataset [16], the images of fog days are from the images of fog days in the DAWN dataset and the real thick fog images collected by ourselves, and the images of rainy days are from a Rain in Driving (RID) set [12]. Based on the above data, we use data augmentation to some images, and the images were re-labeled based on our detection requirements. Finally, we made an automatic driving dataset containing a wide range of real adverse weather, including 5517 images, 1650 in snow, 1962 in fog, and 1905 in rainy weather. The individual categories and their numbers are shown in Table 1. Among them, ordinary vehicles mainly include small sedans SUV, Vans, trucks, buses, coaches, etc. In addition, due to the presence of a very small number of bicycle and motorcycle categories in the image, they are not classified separately and are classified under the category of ordinary vehicles.

Table 1
Type Information of Database.

Type	Car	person	truck
Number	21765	1471	3524

4.2. Experimental Process

The paper first performs road target annotation on the reconstructed data according to the requirements of the paper, and then applies it to the training of BLS and detectors, respectively. When conducting classification training, the labels of the images are divided into three categories: rain, snow, and fog. When training the detector, the image labels are the previously annotated road targets. The next step is to apply the data to the image processing method proposed in our paper to form a new enhanced dataset, and then train the detector for detection to conduct comparative experiments to verify the effectiveness of the framework proposed in our paper. Finally, the models trained with enhanced data are combined to form the final framework.

4.3. Comparison Experiment of BLS Classification

In this part, we adopt the accuracy rate to estimate the ability of the network to classify, and record training time of the network to evaluate the volume of network. Since the final framework needs to be deployed on the edge side, we need lightweight networks, and BLS exactly meets this need. The following Figures 5-6 indicate the improvement of accuracy by BLS incremental learning, and the comparison experiment of accuracy and training time of BLS with other common CNN classification networks. We selected MobileNetV2, which represents CNN lightweight network, AlexNet, which is the classic and effective CNN network, and VggNet16, which is the mainstream classification CNN network.

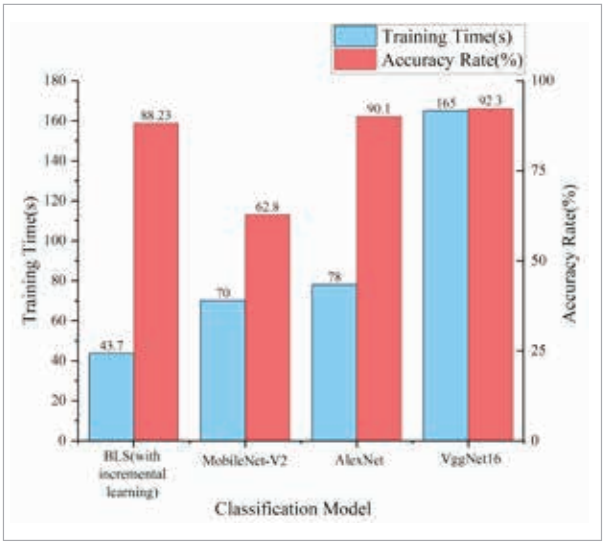
Figure 5

Comparison of detection results.

It can be seen from Figure 6 that incremental learning can increase the classification accuracy of the BLS effectively. In only 43.7 seconds, through 50 rounds of incremental learning, the accuracy of the network classification reached 88.23%, compared with other popular deep learning classification networks, BLS has more advantages than the lightweight network MobileNetV2 in terms of accuracy and training time, Compared with AlexNet and VggNet16, the training time is saved by nearly 2 or even 4 times while the accuracy is only 1.87% and 2.87% lower, respectively.

Figure 6

Comparison of Accuracy Rate and Train Time.



4.4. Detection Result

In this section, we first train the RT-DETR with the pre-augmented dataset and the post-augmented dataset to evaluate whether enhancement based on framework is effective. Figure 5 is comparisons of detection results before and after using REWeather. We use Mean Average Percision (mAP) that represents the average value of each type of AP as our metric, which is widely used to evaluate the performance of popular detectors like YOLO, Faster-RCNN, etc. AP represents the domain encompassed by Precision and Recall. Precision represents the rate of correctly predicted positive samples among all predicted positive samples and only attentions positive samples that makes it distinguishes from accuracy. Precision can be expressed as follows:

$$Precision = \frac{TP}{TP + FP}. \tag{17}$$

Recall can be expressed as follows:

$$Recall = \frac{TP}{TP + FN}, \tag{18}$$

where TP represents number of positive samples correctly discriminated as positive, and FP represents number of negative samples incorrectly

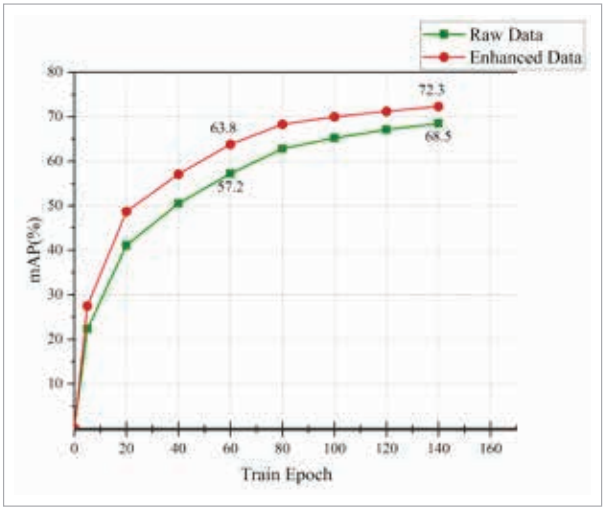
Table 2
Comparison of mAP value of the different kinds of detectors using raw data and enhanced data by our framework.

Methods	Backbone	Database	AP	AP50	AP75	Snow	Fog	Rain
SSD	ResNet50	Raw	31.1	51.9	33.2	63.9	61.8	30.5
SSD	ResNet50	Enabled	29.7	50.9	32.0	62.9	60.4	29.3
Faster-RCNN	ResNet101-FPN	Raw	31.3	57.0	32.9	65.4	67.5	38.2
Faster-RCNN	ResNet101-FPN	Enabled	33.4	59.5	34.7	67.3	70.7	40.9
Yolov5-L	-	Raw	37.1	62.6	-	73.2	73.9	42.1
Yolov5-L	-	Enabled	39.4	65.5	-	74.2	76.3	46.7
Yolov8-L	-	Raw	40.8	65.1	42.5	73.8	76.0	47.8
Yolov8-L	-	Enabled	42.3	68.3	44.3	75.1	78.3	51.7
RT-DETR	ResNet50	Raw	43.9	68.5	46.5	73.7	79.0	53.4
RT-DETR	ResNet50	Enabled	46.4	72.3	49.5	77.9	82.8	57.4

discriminated as positive. FN represents that positive samples are incorrectly discriminated as negative. RT- DETR, YOLOv5, SSD, and Faster-RCNN is trained respectively with raw data and data enhanced by our method to pick out the suitable detectors. The final training comparison results are shown in Table 2.

In order to minimize the model size and reduce deployment difficulty as much as possible, we chose RT-DETR with resnet50 as the backbone network, compared to SSD and Faster RCNN using the optimal backbone network, as well as the most common and latest Yolov5-L and Yolov8-L models in real-time detectors. As seen from Table 2, when the training data is the enhanced data obtained by our method, RT-DETR has the highest mAP among these mainstream detectors, so we choose RT-DETR as the detector in the framework. The improvement in accuracy of RT-DETR also demonstrates the effectiveness of our framework for data augmentation which makes road targets in adverse weather more easily detectable by detectors. Figure 7 shows the change of mAP index of the RT-DETR using raw data and enhanced data along with the training epochs increase. It is evident that after training the model with improved data, mAP of RT-DETR is always better than that trained with original data, and the experiment shows that mAP increased by 3.8 % by using our framework. In addition, we also tested different models based on data from a single adverse weather condition. According to the results in the table, our framework has the best performance under any weather condition.

Figure 7
Comparison of mAP value of the RT-DETR detector using raw data and enhanced data as the training epochs increase.



In addition to comparing the accuracy of the detectors, we also compared the number of parameters and Frames Per Second (FPS) of these detectors in Table 3. Because of the two-stage structure of Faster-RCNN, although the number of parameters is small, the lowest FPS value is only 6 FPS, while SSD using VGG16 backbone network and YOLO-5-L have higher 18 FPS and 22 FPS but larger 138M and 47M parameters, in contrast, RT-DETR has the smallest number of 21M parameters and the highest 47 FPS among these mainstream detectors.

Table 3
Comparison of Detectors.

Detector	Faster-RCNN (ResNet101-FPN)	SSD (ResNet50)	Yolov5-L	Yolov8-L	RT-DETR (ResNet50)
Params(M)	25.6	138	47	43	21
Image/Sec	6	18	22	31	47

4.5. Ablation Study of Framework

To demonstrate the legitimacy of each part of the framework, we used the method of ablation experiment. In Table 4, we used the optimal model trained by the detector using the enhanced data obtained through the framework in this paper as the test model, and proved the rationality and effectiveness of the combination of various parts of our framework by removing each part separately.

Table 4 indicates our framework has the highest mAP value. When Real-ESRGAN enhancement is removed, the mAP value also increases, which shows the effectiveness of guided filtering and single-channel fog removal methods. Meanwhile, when Real-ESRGAN is used for feature enhancement, the mAP value is very limited when guided filtering is used to remove rain to process rainy images instead of dark channel to process foggy images. This is because when foggy images are not directly enhanced to remove fog, the vehicle features are not obvious, which leads to increased fog interference on images after enhancement. On the contrary, when dark channel is used to process foggy images instead of guided filtering to process rainy images, the mAP value is less affected after feature enhancement detection by Real-ESRGAN, because the number of images seriously disturbed by raindrops in rainy day data set is relatively small. Overall, the ablation experiments demonstrate that our framework has the greatest display of abilities and effect of each single method or a combination of methods on the framework.

5. Conclusion

We propose REWeather framework for adverse weather, which has its classification based on BLS, enhanced and denoising based on the fusion of Real-ESRGAN and simple denoising methods, and detection based on RT-DETR. We used non-synthetic dataset to train and test framework and tested our

Table 4
Results of ablation study.

Guided Filter (Rain Removal)	Dark Channel (Fog Removal)	Real-ESRGAN (Enhancement)	Detector	mAP (%)
X	X	X	RT-DETR	68.7
✓	X	X	RT-DETR	68.9
X	✓	X	RT-DETR	69.4
✓	✓	X	RT-DETR	69.6
✓	X	✓	RT-DETR	68.8
X	✓	✓	RT-DETR	69.9
X	X	✓	RT-DETR	69.1
✓	✓	✓	SSD	51.0
✓	✓	✓	Faster-RCNN	56.34
✓	✓	✓	YOLOv5	61.27
✓	✓	✓	**RT-DETR**	**72.3**

framework by deploying it on the edge that is closer to reality. Through experiments, we proved the optimization of our framework for adverse weather automatic driving images, and the experiment shows that our REWeather framework improves the problem of wrong detection and missing detection of autonomous vehicles, elevating the safety of self-driving vehicles in adverse weather. Moreover, how to further reduce the size of the model to achieve a faster and more effective method can be further studied, and we will improve and optimize in the future.

Conflicts of Interest

The authors declared no potential conflicts of interest with respect to the research, author-ship, and publication of this article.

Acknowledgement

This study is supported in part by the Natural Science Foundation of Liaoning Province (grant no. 2024-bs-102), the Research Program of the Liaoning Liaohe Laboratory (Grant no. LLL24KF-01-01), the National Key Research and Development Program of China (Grant No. 2022YFE011400), the Basic Scientific Research Project of the Education Department of Liaoning Province (Grant no. LJ222410142043), JSPS KAKENHI Grant Numbers JP20H04174, JP22K11989, Leading Initiative for Excellent Young Researchers (LEADER), MEXT, Japan, and JST, PRESTO Grant Number JPMJPR21P3, Japan.

Data availability

Data will be made available on reasonable request.

References

1. Carion, N., Massa, F., Synnaeve, G., Usunier, N., Kirillov, A., Zagoruyko, S. End-To-End Object Detection with Transformers. European Conference on Computer Vision. Cham: Springer International Publishing, 2020, 213-229. https://doi.org/10.1007/978-3-030-58452-8_13
2. Chen, C. L. P., Liu, Z. Broad Learning System: An Effective and Efficient Incremental Learning System Without the Need for Deep Architecture. IEEE Transactions on Neural Networks and Learning Systems, 2017, 29(1), 10-24. https://doi.org/10.1109/TNNLS.2017.2716952
3. Chen, Q., Chen, X., Zeng, G., Wang, J. Group Detr: Fast Training Convergence with Decoupled One-To-Many Label Assignment. Arxiv Preprint Arxiv:2207.13085, 2022, 2(3), 12.
4. Dannheim, C., Icking, C., Mäder, M., Sallis, P. Weather Detection in Vehicles by Means of Camera and LIDAR Systems. 2014 Sixth International Conference on Computational Intelligence, Communication Systems and Networks. IEEE, 2014, 186-191. https://doi.org/10.1109/CICSyN.2014.47
5. Fu, X., Huang, J., Ding, X., Liao, Y. Paisley, J. Clearing the Skies: A Deep Network Architecture for Single-Image Rain Removal. IEEE Transactions on Image Processing, 2017, 26(6), 2944-2956. https://doi.org/10.1109/TIP.2017.2691802
6. Hao, H., Zheng, J., Xu, J., Deng, W. Fault Diagnosis Method Based on Principal Component Analysis and Broad Learning System. IEEE Access, 2019, 7, 99263-99272. https://doi.org/10.1109/ACCESS.2019.2929094
7. He, K., Sun, J., Tang, X. Guided Image Filtering. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2012, 35(6), 1397-1409. https://doi.org/10.1109/TPAMI.2012.213
8. He, K., Sun, J., Tang, X. Single Image Haze Removal Using Dark Channel Prior. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2010, 33(12), 2341-2353. https://doi.org/10.1109/TPAMI.2010.168
9. Hodges, C., Bennamoun, M., Rahmani, H. Single Image Dehazing Using Deep Neural Networks. Pattern Recognition Letters, 2019, 128, 70-77. https://doi.org/10.1016/j.patrec.2019.08.013
10. Hu, X., Xu, X., Xiao, Y., Chen, H., He, S., Qin, J., Heng, P.A. Sinet: A Scale-Insensitive Convolutional Neural Network for Fast Vehicle Detection. IEEE Transactions on Intelligent Transportation Systems, 2018, 20(3), 1010-1019. https://doi.org/10.1109/TITS.2018.2838132
11. Kenk, M. A., Hassaballah, M. DAWN: Vehicle Detection in Adverse Weather Nature Dataset. Arxiv Preprint Arxiv:2008.05402, 2020.
12. Ledig C., Theis L., Huszar F., Caballero J., Cunningham A., Acosta A., Aitken A., Tejani A., Totz J., Wang Z., Shi W. Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, 4681-4690. https://doi.org/10.1109/CVPR.2017.19

13. Li S., Araujo I. B., Ren W., Wang Z., Tokuda E. K., Hirata Junior R., Cesar-Junior R., Zhang J., Guo X., Cao X. Single Image Deraining: A Comprehensive Benchmark Analysis. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, 3838-3847. https://doi.org/10.1109/CVPR.2019.00396

14. Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C. Y., Berg, A. C. Ssd: Single Shot Multibox Detector. European Conference on Computer Vision. Cham: Springer International Publishing, 2016, 21-37. https://doi.org/10.1007/978-3-319-46448-0_2

15. Ni, S., Cao, X., Yue, T., Hu, X. Controlling the Rain: From Removal to Rendering. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, 6328-6337. https://doi.org/10.1109/CVPR46437.2021.00626

16. Peng, X., Zhao, N., Xu, L., Bai, S. Vehicle Classification System with Mobile Edge Computing Based on Broad Learning. 2022 IEEE International Conference on Trust, Security and Privacy in Computing and Communications (Trustcom). IEEE, 2022, 1611-1617. https://doi.org/10.1109/TrustCom56396.2022.00231

17. Pitropov, M., Garcia, D. E., Rebello, J., Smart, M., Wang, C., Czarnecki, K., Waslander, S. Canadian Adverse Driving Conditions Dataset. The International Journal of Robotics Research, 2021, 40(4-5), 681-690. https://doi.org/10.1177/0278364920979368

18. Redmon, J., Divvala, S., Girshick, R., Farhadi, A. You Only Look Once: Unified, Real-Time Object Detection. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, 779-788. https://doi.org/10.1109/CVPR.2016.91

19. Ren, S., He, K., Girshick, R., Sun, J. Faster R-Cnn: Towards Real-Time Object Detection with Region Proposal Networks. Advances In Neural Information Processing Systems, 2015, 28.

20. Sakaridis, C., Dai, D., Van Gool, L. Semantic Foggy Scene Understanding with Synthetic Data. International Journal of Computer Vision, 2018, 126(9), 973-992. https://doi.org/10.1007/s11263-018-1072-8

21. Vaswani A., Shazeer N., Parmar N., Uszkoreit J., Jones L., Gomez A. N., Kaiser Ł., Polosukhin I. Attention Is All You Need. Advances In Neural Information Processing Systems, 2017, 30.

22. Wang X., Yu K., Wu S., Gu J., Liu Y., Dong C., Qiao Y., Loy C. C. Esrgan: Enhanced Super-Resolution Generative Adversarial Networks. Proceedings of the European Conference on Computer Vision (ECCV) Workshops, 2018, 0-0. https://doi.org/10.1007/978-3-030-11021-5_5

23. Wang, X., Xie, L., Dong, C., Shan, Y. Real-Esrgan: Training Real-World Blind Super-Resolution with Pure Synthetic Data. Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021, 1905-1914. https://doi.org/10.1109/ICCVW54120.2021.00217

24. Wang, Z., Li, Y., Chen, X., Lim, S. N., Torralba, A., Zhao, H., Wang, S. Detecting Everything in the Open World: Towards Universal Object Detection. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023, 11433-11443. https://doi.org/10.1109/CVPR52729.2023.01100

25. Xu, J., Zhao, W., Liu, P., Tang, X. An Improved Guidance Image Based Method to Remove Rain and Snow in a Single Image. Computer and Information Science, 2012, 5(3), 49. https://doi.org/10.5539/cis.v5n3p49

26. Zang, S., Ding, M., Smith, D., Tyler, P., Rakotoarivelo, T., Kaafar, M. A. The Impact of Adverse Weather Conditions on Autonomous Vehicles: How Rain, Snow, Fog, and Hail Affect the Performance of a Self-Driving Car. IEEE Vehicular Technology Magazine, 2019, 14(2), 103-111. https://doi.org/10.1109/MVT.2019.2892497

27. Zheng, H., Dembélé, S., Wu, Y., Liu, Y., Chen, H., Zhang, Q. A Lightweight Algorithm Capable of Accurately Identifying Forest Fires from UAV Remote Sensing Imagery. Frontiers In forests and Global Change, 2023, 6, 1134942. https://doi.org/10.3389/ffgc.2023.1134942

28. Zhu, X., Su, W., Lu, L., Li, B., Wang, X., Dai, J. Deformable Detr: Deformable Transformers for End-To-End Object Detection. Arxiv Preprint Arxiv:2010.04159, 2020.

29. Zhu, Z., Lei, Y., Qin, Y., Zhu, C., Zhu, Y. IRE: Improved Image Super-Resolution Based on Real-ESRGAN. IEEE Access, 2023, 11, 45334-45348. https://doi.org/10.1109/ACCESS.2023.3256086

30. Zou, Z., Chen, K., Shi, Z., Guo, Y., Ye, J. Object Detection in 20 Years: A Survey. Proceedings of the IEEE, 2023, 111(3), 257-276. https://doi.org/10.1109/JPROC.2023.3238524