



---

*Research article*

## Super-resolution reconstruction algorithm for dim and blurred traffic sign images in complex environments

Yan Ma<sup>1</sup> and Defeng Kong<sup>2,\*</sup>

<sup>1</sup> Wuhan Technical College of Communications, Wuhan, China

<sup>2</sup> School of Mechanical Engineering, Hubei University of Technology, Wuhan, China

\* **Correspondence:** Email: kongdefeng0916@gmail.com; Tel: +18163510917.

**Abstract:** In poor lighting and rainy and foggy bad weather environments, road traffic signs are blurred and have low recognition, etc. A super-resolution reconstruction algorithm for complex lighting and bad weather traffic sign images was proposed. First, a novel attention residual module was designed to incorporate an aggregated feature attention mechanism on the jump connection side of the base residual module so that the deep network can obtain richer detail information; second, a cross-layer jump connection feature fusion mechanism was adopted to enhance the flow of information across layers as well as to prevent the problem of gradient disappearance of the deep network to enhance the reconstruction of the edge detail information; and lastly, a positive-inverse dual-channel sub-pixel convolutional up-sampling method was designed to reconstruct super-resolution images to obtain better pixel and spatial information expression. The evaluation model was trained on the Chinese traffic sign dataset in a natural scene, and when the scaling factor is 4, the average values of PSNR and SSIM are improved by 0.031 when compared with the latest release of the deep learning-based super-resolution reconstruction algorithm for single-frame images, MICU (Multi-level Information Compensation and U-net), the average values of PSNR and SSIM are improved by 0.031 dB and 0.083, and the actual test average reaches 20.946 dB and 0.656. The experimental results show that the reconstructed image quality of this paper's algorithm is better than the mainstream algorithms of comparison in terms of objective indexes and subjective feelings. The super-resolution reconstructed image has a higher peak signal-to-noise ratio and perceptual similarity. It can provide certain technical support for the research of safe driving assistive devices in natural scenes under multi-temporal varying illumination conditions and bad weather.

**Keywords:** traffic sign images; super-resolution reconstruction; residual module; multi-layer feature

fusion; sub-pixel convolution

**Mathematics Subject Classification:** 68T45

---

## 1. Introduction

Safe driving is a core technical requirement of motor vehicle driving aids, and, to ensure a high driving safety coefficient, it is necessary to accurately identify traffic signs in the road environment, which are a core technical support for a safe driving aid systems [1]. However, due to the complexity of the traffic road environment and the large differences in visibility of traffic signs at different times of the day and in different weather environments, such as dim morning, evening, and night time, midday with strong illumination, and rainy, hazy and foggy days with blurred vision, traffic sign images show degradation phenomena such as dim and blurred and loss of information on the edges of the signs [2], and the lack of detailed information leads to low spatial resolution. Therefore, it is necessary to study the super-resolution reconstruction technology of traffic sign images at different times of day and in different weather environments, and use the image super-resolution reconstruction technology to improve the image quality of traffic signs, make the algorithm easier to deploy in safe driving assistance devices, and promote the intelligent development of driving technology [3].

Single image super-resolution reconstruction (SISR) is one of the most fundamental image processing problems in computer vision [4], the task of which is to reconstruct one or more low-resolution images (LR) into high-resolution images (HR), and because there are several factors that degrade an HR image into an LR image, image super-resolution reconstruction is a challenging subject [5]. This technique has a wide range of applications in fields such as medical imaging [6], satellite remote sensing [7], and security [8]. Early super-resolution reconstruction methods usually used nearest neighbor interpolation, bilinear interpolation, and bicubic interpolation [9], and although the interpolation method has small computational complexity and high reconstruction efficiency, the method reconstructs high-resolution images from local pixel information, which is not ideal when the magnification is large. Reconstruction-based methods appeared later [10], which are based on the image degradation model, and through the extraction of key information in the low-resolution image combined with a priori knowledge of the gradient and edge of the unknown super-resolution image, the mapping relationship from the LR image to the HR image is constructed and the loss function solution to obtain the high-resolution image is optimized [11]. Reconstruction-based methods have better reconstruction results compared to interpolation. Still, the model is inefficient, the reconstruction effect is easily affected by the accuracy of the regularization parameters, resulting in the texture details not being effectively recovered, and the computational time is greater, which makes it difficult to take into account real-time needs. Learning-based methods usually utilize a large number of LR images to train the model, from which some kind of implicit mapping relationship between LR images and HR images is learned, and then the LR images are reconstructed into the corresponding high-resolution images according to the learned implicit relationship, such as in sparse coding [12], stream learning [13], dictionary learning [14], etc. Although the learning-based methods have achieved further improvement, the optimization and updating process of network parameters is very complicated, and the model is usually not able to recover the texture details effectively. The optimization and updating process is very complicated, and the model is usually difficult to fit into the best target, leading to unstable reconstruction results.

In recent years, deep learning has developed rapidly in the field of image super-resolution reconstruction, and a variety of SR methods based on deep learning have appeared to solve the various problems existing in traditional algorithms. Dong et al. [15] pioneered the introduction of convolutional neural networks for the task of image super-resolution reconstruction, and designed a super-resolution convolutional neural network (SRCNN) that has only three layers to realize feature extraction, feature nonlinear mapping, and reconstruction. SRCNNs achieve good end-to-end reconstruction results and lay the foundation for deep learning in the field of image super-resolution reconstruction. Although SRCNNs shows good reconstruction performance compared to traditional SR methods, the SRCNN algorithm has fewer layers in the network structure, which makes it difficult to fully utilize the image context information resulting in HR image clarity that needs to be further improved. To solve the problem of incomplete feature extraction caused by the small number of network layers, Kim et al. [16] used the ResNet module to construct the ultra-deep SISR network model (VDSR). The VDSR network utilizes the idea of residual learning, and learns the high-frequency information between low- and high-resolution images by jumping connections, which improves the training efficiency and the clarity of the reconstruction of the texture details; however, the deepening of the network also brings about a drastic increase in the number of parameters, which leads to the improvement of the clarity of the HR image. Tai et al. [17] combined ResNet and RNN(Recurrent Neural Network) to propose a deep recurrent residual network (DRRN) for image super-resolution reconstruction, which improves the performance of the network without introducing too many parameters by adopting recursive learning for some of the convolutional layers. Although the parameter-sharing mechanism of the RNN can enable the network model to maintain a small number of parameters while learning higher-level features, it still cannot avoid the problems of gradient vanishing and redundant information training consumption faced by ultra-deep networks. Zhang et al. [18] applied the channel attention mechanism to image SR for the first time in 2018, and combined it with the residual module to propose the residual channel attention network (RCAN), although the addition of the attention mechanism makes the network learn more important edge detail features, the authors simply interspersed the plug-and-play channel attention module in one layer of the network, which makes the reinforcement of important features limited. The excellent performance of the attention mechanism and residual learning became an important part of the later excellent single-frame image superpixel reconstruction (SISR) network models; for example, the second-order attention network (SAN) proposed by Dai et al. [19] introduces the second-order channel attention mechanism, which has better reconstruction performance on drastically varied and diverse targets, but the channel attention mechanism overemphasizes the importance of the channel information and neglects the spatial relationship between pixels, thus lacking a detailed description of texture features. Wei et al. [20] proposed the component partition network model (CDC) to explore the importance of different components by constructing three component attention blocks related to planes, edges, and corners to solve the superpixel reconstruction problem with the idea of divide-and-conquer, but the components independent of each other impede the interactive learning of spatial features of the channels, which makes the reconstructed image incongruous in terms of light and dark, and jumping edge pixels. As deep learning continues to develop in the field of image super-resolution reconstruction, some researchers have recently proposed reconstruction models with better performance, such as Zhang et al. [21] who proposed a cascaded visual attention network (CVANet) for single-image super-resolution reconstruction to solve the problem of the deep convolutional neural network utilizing feature maps, channels, and insufficient pixels, but the independent reinforcement design strategy

adopted makes the feature maps, channels, and pixels still lack information exchange among each other, leading to problems such as loss of image details and pixel distortion. Wang et al. [22] designed a Transformer-based Terrain Neural Network (TTSR) for the super-resolution reconstruction of digital elevation models by introducing the Transformer structure, and, compared with the traditional method, the method's elevation accuracy, slope accuracy, and root-mean-square error (RMSE) of slope accuracy are reduced by about 6%–30%, 4%–16%, and 1%–9%, respectively. Although progress is made in performance, the transformer structure drags down the model training speed and fitting efficiency considerably, and is difficult to deploy in common training platforms as well as edge computing devices, which affects the practical application effect. Chen et al. [23] used the multilevel information through the multi-level compensation and U-net (MICU) to realize image super-resolution reconstruction, but the U-net structure adopts the strategy of down-sampling and up-sampling, which is less effective for the reconstruction of larger targets.

However, although there are many newly proposed image reconstruction methods, there are fewer studies related to the reconstruction of traffic sign images, which is a research shortcoming in the development of intelligent assistive devices for safe driving, mainly adopting traditional image processing techniques and classical machine learning methods, such as Qu [24] who proposed an Adaboost integration algorithm based on the image keypoint statistical transform (MCT) features to realize the classification and recognition of traffic signs under complex lighting conditions; Zhang [25] and others used BP neural networks for real-time recognition of speed limit traffic signs. Although traditional image processing technology and machine learning methods have obtained certain recognition accuracy, because the environment where the traffic signs are simple, the difficulty of recognition is not a big problem. In addition, traditional methods have not high accuracy, complicated algorithm design, poor operability, and tedious deployment difficulty. Xu et al. [26] used the improved Cascade R-CNN depth model to recognize traffic signs in rain, snow, fog, and other inclement weather, and achieved better recognition accuracy; but, this higher recognition accuracy is based on short-horizon scenes and environments with better visual conditions, so for long horizon scenes at different hours of illumination and inclement weather environments, dim, fuzzy, and unclear edges of the traffic signs do not have very good generalization performance and recognition accuracy. However, a high-performance complex system for safe driving must take into account the accurate recognition of traffic signs in a long field of view, and it is important for safe driving to accurately recognize traffic signs within a sufficiently safe distance [27]. Since long field-of-view traffic signs usually have smaller targets, fuzzy and dim traffic sign bodies, unclear edges of sign prompt messages, and lower resolution, super-pixel reconstruction of low-resolution traffic signs is needed to obtain larger visible images and clear sign edges, which can help safe driving assistance systems to recognize traffic signs in advance in longer field-of-view ranges and to make safe driving warnings [28]. From the above review of image super-resolution reconstruction methods, it can be understood that, although the attention mechanism and residual learning have good extraction effect on LR image texture, details, and edge high-frequency information, the jump connection of the residuals also passes a large amount of low-frequency information to the high-level feature layer, which affects the reconstruction effect of the reconstruction layer on the texture details; in addition, the interpolation and the inverse convolution upsampling methods form discontinuous jagged edges. While sub-pixel convolution can alleviate this problem, it still cannot eliminate the obvious artificial traces.

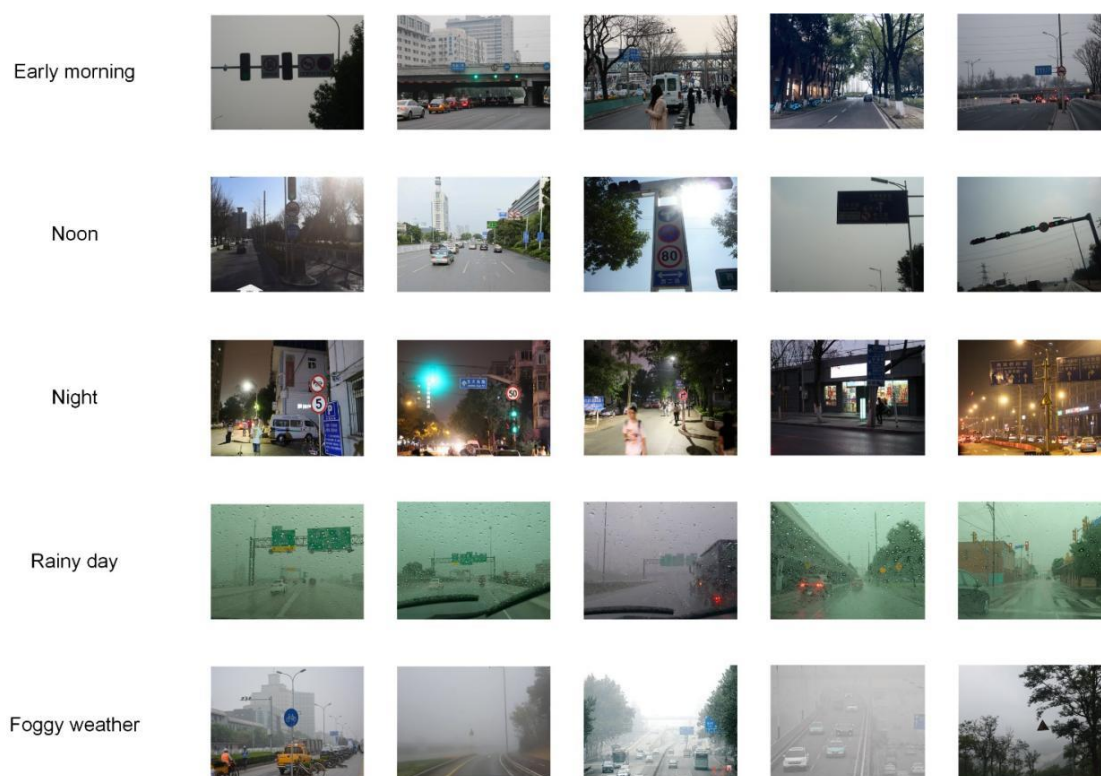
Based on the above problems, this paper proposes a super-resolution reconstruction method for low-resolution traffic sign images by integrating attention residuals, an important feature fusion

strategy, and double inverse channel sub-pixel convolution. To do so, first the channel spatial hybrid attention mechanism (CBAM) [29] is fused in the jump connection side of the residual module so that the network learns the relationship between channels and spatial location information in the shallow image at the same time, the channel attention mechanism optimizes the contribution weights between image channels to give better representation to important channels, and the spatial attention mechanism parses high and low-frequency information of different spatial locations within the same channel to extract important high-frequency information and pay more attention to it. When learning the image channel and spatial information, the network can adaptively extract important features in the feature map according to their importance and ignore a large amount of redundant low-frequency information to solve the problem of unclear details and texture of dim and blurred traffic sign images. Second, the color, detail, texture, spatial, and semantic features acquired by the different layers can be adequately fused so that different feature layers can capture the color and detail of the image. The different feature layers can capture different information so that the model has better feature representation and robustness, and, at the same time, by extracting representative features, the impact of noise and redundant information on the network model processing can be reduced. Finally, for the problem of pixel discontinuity and jumping traces in the reconstruction process, a double inverse channel sub-pixel convolution method is proposed. Sub-pixel convolution up-sampling usually adopts the forward channel to arrange the spatial information in the positive order to expand the image size and to increase the spatial information, the original feature map channel is reversed and sampled to arrange the spatial information in the same way, and then the two times of exactly inverse information is fused globally to eliminate obvious artifacts and the pixel jumping phenomenon, and to improve the reconstruction capability of the traffic sign image. For this paper, the main contributions are as follows:

- (1) We propose an attention residual network for backbone feature extraction, which improves the extraction ability of edge and texture features by strengthening important detail features and filtering low-frequency redundant information;
- (2) We design a hierarchical feature fusion method to fully fuse the features extracted from different feature layers and learn the intrinsic mapping relationship between different layers to improve the model generalization ability and robustness;
- (3) We construct a dual inverse channel sub-pixel convolutional upsampling structure to increase the spatial information content, enhance global feature fusion, and eliminate artificial traces and pixel-jagged discontinuities.

Most of the current deep learning-based single-frame image super-resolution reconstruction algorithms are trained on a large scale based on public datasets, such as ImageNet [30], OST [31], DPED [32], etc., which mainly consist of common objects in life, such as people, animals, plants, transportation, buildings, landscapes, and other subjects. Since learning algorithms will have different performances for specific learning objects, the image hyperpixel reconstruction algorithm based on public datasets does not reconstruct well when facing dim and blurred traffic sign images. To address this problem, we first construct a real traffic sign dataset in a natural environment, and then design an adapted deep network model for feature extraction and image reconstruction. Therefore, in this paper, we first construct a dataset of traffic sign images in a natural environment with multiple hours of illumination and severe weather for network model training, which is rich in scenes and fully adapted to the real traffic scene so that the model has a better generalization ability. The traffic sign dataset is collected from natural scene cameras or Baidu Street View under different times, weather conditions, lighting conditions, and motion blur conditions. The types of traffic signs are mainly traffic signs and

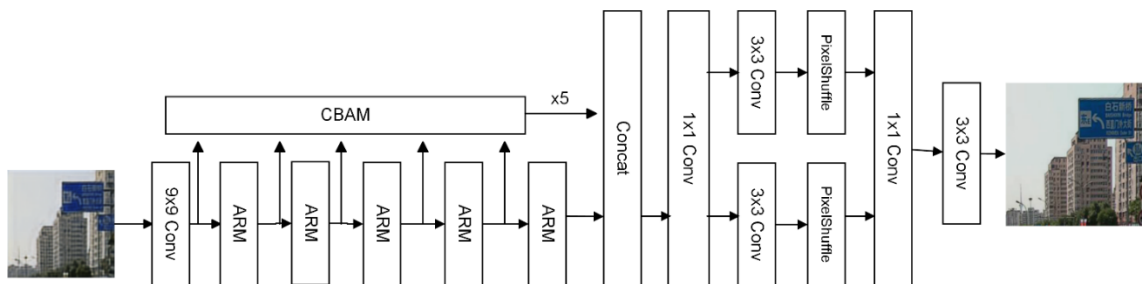
traffic panels; the locations of the collected images include road intersections, accident-prone road sections, and vehicle-only road locations; the periods of the collected images include the early morning when the sky is white, the middle of the day when the light intensity is high, dusk time, and night time, and other all-weather periods; the weather conditions include sunny, foggy, and rainy days, damaged images and extremely fuzzy images are eliminated, and the traffic sign data set is generalized to natural scenes. Selected valid images, the images are uniformly cropped into  $640 \times 480$  size images, and finally 5000 traffic sign images are obtained which reflect the real traffic scenes in the real natural environment and have a certain degree of representativeness. Figure 1 shows examples of dataset images. It is difficult for existing devices to acquire paired LR images and HR images in the same scene, and, usually, only HR images are acquired, and then the corresponding LR images are obtained by the mathematical degradation model. However, the actual LR images are affected by blurring, noise, downsampling, image compression, and many other unknowable and complex factors. Since LR images obtained using the same mathematical degradation model are structurally similar, they tend to obtain higher reconstruction metrics when composing a test set. To avoid such spurious reconstruction performance guides, the training set and the test set will be synthesized into a specific dataset using different degradation model constructions. In this paper, simple downsampling will be used to obtain the training set LR images, and downsampling with a combination of blurring and noise will be used to obtain the test set LR images.



**Figure 1.** Partial images of the image dataset of traffic signs in five environments: early morning, midday, night, rainy day, and foggy day.

## 2. The network model of this paper

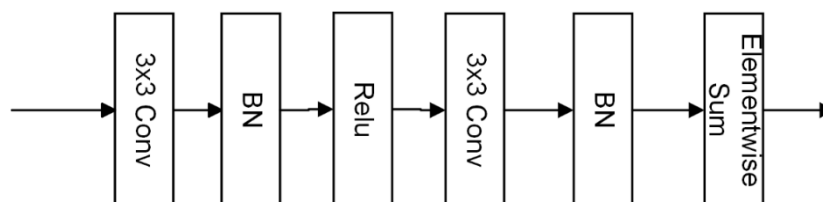
In this paper, we propose a super-resolution reconstruction network model for low-resolution traffic sign images that integrates the attention residual, feature fusion mechanism, and double inverse channel sub-pixel convolution, and the complete network model consists of 3 parts: 1) attention residual module; 2) feature fusion mechanism; 3) image reconstruction module, and the network structure is shown in Figure 2. First of all, the convolution, normalization, and activation functions constitute the basic feature extraction units for cyclic stacking, and the attention mechanism and residual learning ideas are used to link the basic feature extraction units. Then, the fusion layer receives feature information from different depth feature extraction layers, superimposes them on the channel dimension, performs feature fusion and compresses the channel dimension, fully combines the detailed texture features and semantic features of different levels of feature maps, and improves the quality of the reconstructed image. Finally, the image reconstruction layer designed a novel dual inverse channel sub-pixel convolution algorithm to realize the up-sampling of the image, and used the convolution layer for pixel adjustment to complete the final reconstructed image.



**Figure 2.** The overall network structure of image super-resolution reconstruction model.

### 2.1. Attention residual module (ARM)

In the feature extraction stage, the basic feature extraction module is composed of three functional blocks of a  $3 \times 3$  convolutional layer, normalization layer, and Relu activation function layer sequentially arranged and repeated once to constitute the basic feature extraction unit. The network structure is shown in Figure 3, and the logical computational relationship of this module can be expressed by Eq (1).



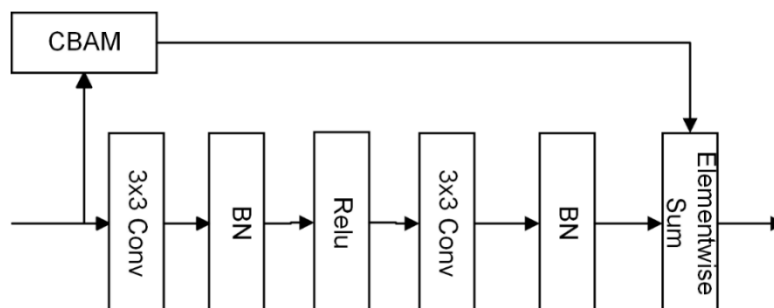
**Figure 3.** Basic feature extraction module.

$$T = \delta \left( BN \left( f_{3 \times 3} \left( I_{LR} \right) \right) \right), \quad (1)$$

where  $I_{LR}$  denotes a low-resolution image,  $f_{3 \times 3}$  denotes a convolutional layer with a convolutional kernel size of 3, BN denotes a normalization layer,  $\delta(\cdot)$  denotes an activation function, where the ReLU activation function is used here, and  $T$  denotes the extracted feature vector. Combining the attention residual structure with the basic feature extraction module constitutes a new type of feature extraction module with residual mapping function and attention to the important features of attention, and the structure of the backbone feature extraction network stacked with this new type of module is shown in Figure 4. Dim and fuzzy traffic signs usually contain rich low-frequency information. Still, the edge detail mainly exists in the high-frequency information, while the high-frequency information is weak. The detail features are gradually lost in the process of continuous convolution, so it is necessary to introduce the residual jump-connection branch, which adds shallow high-frequency information after a certain stage of network learning and strengthens the expression of the edge detail features. However, the jump-connection of the residual structure strengthens the deeper transfer of high-frequency information, and it is more of a color change. Although the residual structure of jump connections enhanced the transfer of high-frequency information to the deeper levels, this process transferred information with more low-frequency information such as color, brightness, etc., and the low-frequency information had a dissociative effect on the deeper high-level semantic information, which is not conducive to the accurate recognition of the target, the addition of the attention mechanism to the branch of the jump connection can effectively attenuate the transmission of the redundant information, the attention mechanism [29,33,34] can give more weight to the important information through the learning features to strengthen the expression and transmission of the important features, and the computational expression is shown in the Eq (2).

$$T_n = BN \left( f_{3 \times 3} \left( \delta \left( BN \left( f_{3 \times 3} \left( T_{n-1} \right) \right) \right) \right) \right) + A_{CBAM} \left( T_{n-1} \right), \quad (2)$$

where the feature map  $T_1$  is not computed because the first feature layer does not apply to Eq (2), where  $n \geq 2$  in Eq (2),  $T_{n-1}$  is the input feature of the  $n$ th attention residual module,  $f_{3 \times 3}$  denotes the convolutional layer with convolutional kernel size 3, BN denotes the normalization layer,  $\delta(\cdot)$  denotes the ReLU activation function,  $A_{CBAM}(\cdot)$  denotes the spatial hybrid attention mechanism of the CBAM channel, and  $T_n$  denotes the  $n$ th attention residual output feature of the module. According to Figure 4 and Eq (2), the pseudo-code algorithm is as Algorithm 1.



**Figure 4.** Attention residual module.



**Algorithm 1:** Attention residual module feature extraction algorithm

---

```

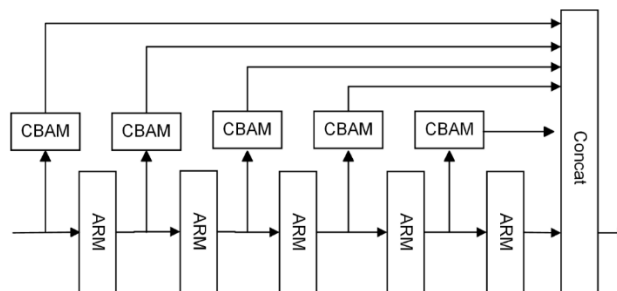
1   function ARM( x );
      Input: Original low-resolution image
      Output: eigenmaps
2   if x = 0 then
3       return 0;
4   else
5        $x_1 = A_{CBAM}(x)$ ;
6        $x_2 = f_{3 \times 3}(x)$ ;
7        $x_3 = \text{BN}(x_2)$ ;
8        $x_4 = \delta(x_3)$ ;
9        $x_5 = f_{3 \times 3}(x_4)$ ;
10       $x_6 = \text{BN}(x_5)$ ;
11       $x = x_1 + x_6$ ;
12      return x;
13  end

```

---

## 2.2. Feature fusion mechanisms

Feature fusion is the fusion of shallow network features and deep network features. The shallow network has covariant features required for localization, which is crucial for the accurate location of objects in the reconstructed image, and the deep network has more high-level semantic features to distinguish objects, which ensure the consistency of the content of the reconstructed image during the image reconstruction process. Thus, the shallow and deep features are indispensable for the reconstruction of high-resolution images. Each feature layer has its special and important feature information, so it needs to be extracted independently to interact with other independent feature layers in the feature fusion session to enrich the details and semantic features of the image reconstruction layer and improve the quality of image reconstruction. As the feature extraction method learns the jump connection of the residual network, there will still be a large amount of redundant information transfer, so it will be integrated into the channel space attention mechanism module (CBAM) in each connection branch, attenuating the role of invalid information to discrete effective features. The network structure diagram of the combination of the feature extraction module and the feature fusion mechanism is shown in Figure 5.



**Figure 5.** Hierarchical important feature fusion strategy.

In this paper, the design of feature fusion mechanism is based on the jump connection of different layers of attention residual modules and residual learning, the LR image input into the network firstly do the convolutional kernel size of 9 for large scale shallow feature extraction, and stretch the first feature layer channel into 64 channels; then, after the first feature layer successively stack five attention residual modules (ARM), and elicit the input feature layer of different modules with the last one module output feature layer Concat operation; finally,  $3 \times 3$  convolution is used to do information fusion on the last Concat feature layer and output it into 64 channels to provide rich spatial and pixel features for the subsequent reconstruction layer, and the whole process can be expressed by Eqs (3) and (4).

$$F_1 = A_{CBAM} \left( \sigma \left( f_{9 \times 9} (I_{LR}) \right) \right), \quad (3)$$

where  $I_{LR}$  denotes the original low-resolution image,  $f_{9 \times 9}$  denotes the convolution operation with convolution kernel size 9,  $\delta(\cdot)$  denotes the Relu activation function, and  $A_{CBAM}(\cdot)$  denotes the CBAM channel space hybrid attention mechanism, where the low-resolution original image passes through the shallow feature extraction network and is weighted by the attention mechanism to obtain the feature map  $F_1$ .

$$F_n = A_{CBAM} (T_n), \quad (4)$$

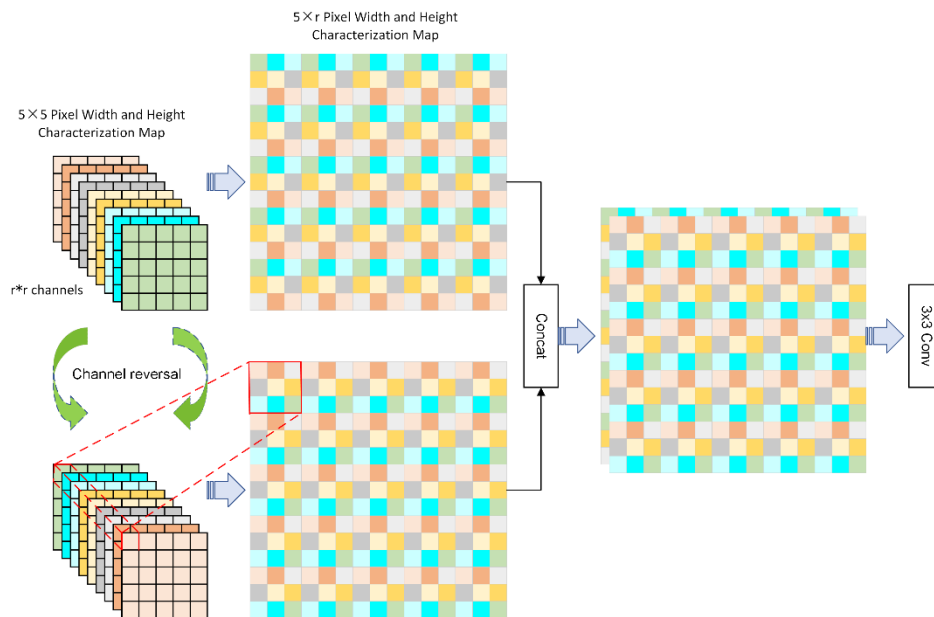
where  $n \geq 2$  is the logical equation for calculating the feature layers  $F_2 - F_6$ ,  $T_n$  is described in Eq (2), and Eq (4) represents the new feature maps obtained by attentional weighting of the  $T_2 - T_6$  feature maps.

$$T_M = \sigma_{Leaky-Relu} \left( f_{3 \times 3} \left( Concat [F_1, F_2, F_3, F_4, F_5, F_6] \right) \right). \quad (5)$$

The operation logic to be expressed in Eq (5) is to superimpose different levels of feature layers in the channel dimensions and then apply convolutional operations to adjust the number of channels of the fused feature layers and learn different levels of features, and finally obtain the new feature layers through the activation function.  $F_n$ ,  $n = 1, 2, 3, 4, 5, 6$ , denotes the feature layer after different stages of extraction and attention weighting,  $Concat[\cdot]$  is the superposition of different layers of feature maps in the channel dimension,  $f_{3 \times 3}$  denotes the convolution operation with a convolution kernel size of 3,  $\delta_{Leaky-Relu}(\cdot)$  denotes the Leaky-Relu activation function [35], and we set  $\alpha = 0.2$  to prevent negative output necrosis.

### 2.3. Image reconstruction module

The reconstruction algorithm proposed in this paper is a double inverse channel sub-pixel convolution method, as shown in Figure 6. The core idea is to carry out two channels for inverse sub-pixel convolution up-sampling to obtain two high-resolution images with the same dimensions and depths, but with opposite elements in the unit space of the feature map, and then superimpose two high-resolution images in channel dimension, and the final image of the required magnification will be obtained through a  $3 \times 3$  convolution layer, which is calculated as follows:



**Figure 6.** Dual inverse channel sub-pixel convolution algorithm.

$$I_{Up}^+ = f_{Up}(F_M^+), \quad (6)$$

where  $F_M^+$  denotes the feature map of the channel in forward order,  $f_{Up}$  denotes the subpixel convolutional upsampling operation, and  $I_{Up}^+$  denotes the high-resolution feature map obtained from the forward-ordered channel feature map by the subpixel convolutional operation.

$$I_{Up}^- = f_{Up}(F_M^-). \quad (7)$$

In Eq (7),  $F_M^-$  denotes the feature layer of the channel in reverse order,  $f_{Up}$  is the same as in Eq (6), and  $I_{Up}^-$  denotes the high-resolution feature map obtained from the reverse-order channel feature map by sub-pixel convolution operation.

$$I_{SR} = f_{3 \times 3}(\text{Concat}[I_{Up}^+, I_{Up}^-]). \quad (8)$$

In Eq (8),  $\text{Concat}[\cdot]$  denotes the positive and negative high-resolution feature maps superimposed in the channel dimension, and  $f_{3 \times 3}$  denotes the convolution operation with a convolution kernel size of 3. According to Eqs (6)–(8), the steps to obtain the super-resolution image are as follows: First, obtain the high-resolution feature maps  $I_{Up}^+$  and  $I_{Up}^-$  of the forward and inverse channel feature layers, respectively, and then superimpose the two types of feature maps  $I_{Up}^+$  and  $I_{Up}^-$  in the channel dimension. Then, use the  $f_{3 \times 3}$  convolution to organize the fused high pixel feature maps to obtain the final reconstructed high-resolution image. The pseudo code of this algorithm is in Algorithm 2.

**Algorithm 2:** Dual inverse channel subpixel convolution algorithm

---

```

1  function DICSC( x );
   Input: Deep feature maps
   Output: Dual inverse channel fusion feature map
2  if x = 0 then
3      return 0;
4  else
5      if x =  $F_M^+$  then
6           $I_{Up}^+ = f_{Up}(x)$ ;
7      end
8      if x =  $F_M^-$  then
9           $I_{Up}^- = f_{Up}(x)$ ;
10     end
11      $x_1 = \text{Concat}[I_{Up}^+, I_{Up}^-]$ 
12      $x_2 = f_{3 \times 3}(x_1)$ 
13     return  $x_2$ ;
14 end

```

---

#### 2.4. Loss function

The L1-paradigm loss function is used to constrain the model at the pixel level to make the reconstructed high-resolution image as close as possible to the real high-resolution image by calculating the error of pixel values at the corresponding pixel positions in the reconstructed high-resolution image (SR) and the real high-resolution image (HR) [36], and, due to the good performance of the L1-paradigm loss function in the description of the details, most of the image hyper pixel algorithms nowadays use the L1-paradigm loss function to guide the model training. The L1-paradigm loss function is used to guide the model training, so this paper also adopts the L1-paradigm loss function to optimize the network parameters. For a given training set of traffic sign images,  $\{I_{LR}^i, I_{HR}^i\}_{i=1}^M$  contains M pairs of low and high-resolution images, and the specific expression of the loss function is

$$L_1(\theta) = \frac{1}{M} \sum_{i=1}^M \|I_{HR}^i - I_{SR}^i\|_1, \quad (9)$$

where  $\theta$  denotes the set of parameters to be learned in the network  $\theta = (\omega_1, b_1; \omega_2, b_2; \dots; \omega_n, b_n)$ , including the weights  $\omega_i$  and bias  $b_i$  of each network layer, HR denotes the  $i$ th real high-resolution image, and  $I_{SR}^i$  denotes the  $i$ th super-resolution image after the algorithm reconstruction is completed.

### 3. Experimental setup and interpretation of evaluation indicators

#### 3.1. Experimental platform and parameter settings

The experimental hardware environment for the model training test in this paper is as follows: the CPU of the computer is a 12th Gen Intel® Core™ i5-12600KF 3.70GHz, the system memory is 16

GB, and the graphics card is an NVIDIA GeForce RTX 3070 GPU with 8 GB of video memory capacity. The software environment is the Windows 10 operating system, Pycharm compilation environment, PyTorch1.12 deep learning framework, CUDA 11.6 accelerated computing platform, Anaconda 3.0 environment manager, and the programming language is Python 3.8. After many experimental explorations, the in-depth traffic sign image hyperpixel reconstruction network model proposed in this paper and important hyperparameters are summarized, and we find that the model achieves stable and reliable performance when the model parameters are set as in Table 1.

**Table 1.** Experimental setup.

Set item	parameter
Iteration	200
Batch size	4
Initial learning rate	2e-4
Min learning rate	(2e-4)*0.01
Optimizer	Adam
momentum	0.9
Weight decay	0
Learning rate decay type	COS
thread	4
Up-sample multiple	2/4

### 3.2. Objective evaluation metrics for image super-resolution reconstruction algorithms

Objective evaluation metrics assess image quality through mathematical models and algorithms, which have the advantages of simplicity, efficiency, and reflecting the real phenomenon, among which the full-reference type evaluation metrics Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity (SSIM) are the most popular objective evaluation metrics being used [37]. Therefore, this paper experimentally adopts the PSNR and SSIM metrics to measure the reconstruction quality of traffic sign images.

#### 3.2.1. Peak signal-to-noise ratio

PSNR is an important image evaluation metric that evaluates the degree of distortion of the reconstructed image at the pixel level by calculating the error between the pixel values of the corresponding pixel positions of the reconstructed SR image and the real HR image. PSNR is measured in dB, and the value of PSNR is directly proportional to the quality of the reconstructed image. The larger the value the better the reconstructed quality, and it is calculated as

$$PSNR = 10 \log_{10} \left( \frac{MAX_I^2}{MSE} \right), \quad (10)$$

where  $MAX_I$  is the maximum pixel value of the real image, and  $MSE$  represents the average value of the energy of the difference between the real image and the noise image.

### 3.2.2. Structural similarity

SSIM is also an important evaluation index of image super-resolution reconstruction performance, which is based on the existence of strong correlations in natural images. These correlations carry important information about the structure of the object in the human visual scene by detecting whether the structural information is altered or not to perceive the approximation information of the image distortion and to measure the similarity between the two images. SSIM is mainly composed of the information in three parts: luminance, contrast, and structure between the images. The SSIM measures the similarity between two images. The metric is more in line with the human visual system, and the closer the value is to 1 the better the image reconstruction performance.

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\mu_\sigma^2 + \mu_\sigma^2 + c_2)}, \quad (11)$$

where  $x, y$  represent the real traffic sign image and the reconstructed traffic sign image by the algorithm, respectively;  $\mu_x, \mu_y$  denote the average gray scale value of the real and reconstructed traffic sign images, respectively;  $\delta_{xy}$  denotes the covariance of  $x$  and  $y$ ;  $\delta_x$  represents the variance of the real image;  $\delta_y$  represents the variance of the reconstructed image;  $c_1 = (k_1L)^2$ ,  $c_2 = (k_2L)^2$  are constants used to maintain stability;  $L$  is the dynamic range of the pixel values;  $k_1 = 0.01$ ; and  $k_2 = 0.03$ . The mean is used as an estimate of brightness, the standard deviation as an estimate of contrast, and the covariance as a measure of structural similarity.

## 4. Experimental results and analysis

### 4.1. Analysis of objective indicators

To verify the performance of the algorithms in this paper, mainstream representative image super-resolution reconstruction algorithms are selected to do comparison experiments, which contain the traditional classical bicubic interpolation method Bicubic [38] and the early excellent deep learning methods SRCNN, VDSR, DRRN, RCAN, and CDC mentioned in the review section, as well as the newly released deep learning based reconstruction algorithms CVA, TTSR, MICU, etc. which are trained to the same experimental subjects under the same experimental conditions, and performance comparison experiments with 2x and 4x downsampling rates are conducted on the test set of different periods at early morning, noon, and night, and different bad weather scenarios on rainy and foggy days. Table 2 records the comparison results of this paper's image reconstruction algorithm with other representative algorithms on the image reconstruction quality evaluation metrics, PSNR and SSIM, on the test set of five types of scenarios that lead to dim and blurred traffic signs. Observing the results in Table 2, it can be found that the super-resolution reconstruction performance PSNR and SSIM metrics of image reconstruction are better than other algorithms at different downsampling magnification tests. Specifically, when the magnification is small, the advantage of the algorithm proposed in this paper over other algorithms is not obvious, due to the 2 times downsampling rate of the LR image still retaining a large amount of pixel structure information, that is, the image is not dimmed and blurred much, this means that super-resolution reconstruction of LR images with 2-fold downsampling is less

difficult, while reconstruction of LR images with 4-fold downsampling is more difficult. At this time, this paper algorithm to test the performance of super-resolution reconstruction of the obtained PSNR and SSIM indicators are significantly higher than other algorithms, indicating that the image super-pixel reconstruction algorithm proposed in this paper has obvious advantages in the performance of reconstruction of traffic sign images with worse visibility.

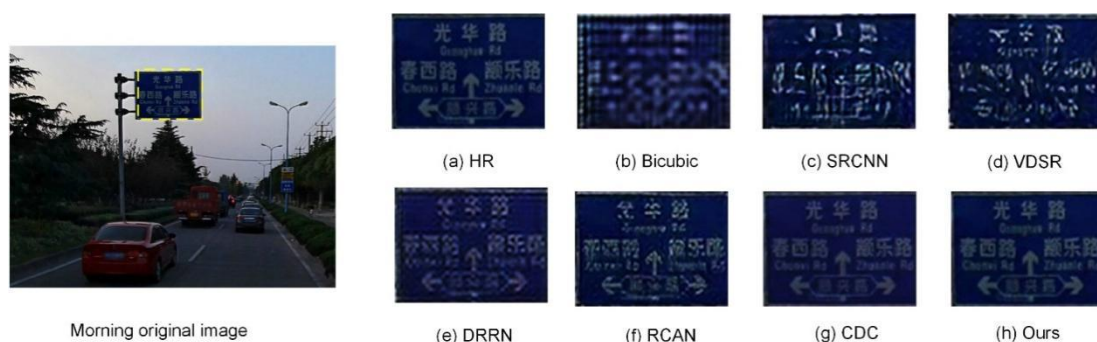
**Table 2.** Comparison of PSNR and SSIM of different algorithms for five environments with traffic sign test set scaling factor of 2/4 times.

Measure	Model	Morning	Noon	Night	Rain	Fog	Mean
		PSNR/SSIM	PSNR/SSIM	PSNR/SSIM	PSNR/SSIM	PSNR/SSIM	PSNR/SSIM
x 2	Bicubic	19.254/0.684	16.548/0.473	23.821/0.617	23.168/0.627	20.397/0.358	20.567/0.557
	SRCNN	21.041/0.722	18.528/0.597	25.281/0.703	24.906/0.682	22.657/0.444	22.483/0.650
	VDSR	22.384/0.718	19.954/0.624	25.328/0.718	25.286/0.704	22.914/0.483	23.164/0.679
	DRRN	22.594/0.701	20.185/0.683	25.89/0.724	25.824/0.714	23.345/0.516	23.964/0.686
	RCAN	22.867/0.716	20.675/0.709	25.973/0.728	26.627/0.721	23.728/0.594	24.428/0.701
	CDC	23.673/0.720	21.159/0.718	26.004/0.730	26.991/0.728	24.297/0.602	24.872/0.704
	CVA	23.743/0.717	22.312/0.719	26.174/0.728	27.094/0.654	24.341/0.602	24.733/0.684
	TTSR	23.792/0.721	22.427/0.720	26.206/0.731	27.137/0.696	24.479/0.604	24.808/0.694
	MICU	24.291/0.722	22.513/0.726	26.297/0.733	27.208/0.712	24.608/0.605	24.983/0.700
	Ours	24.346/0.724	22.685/0.730	26.311/0.739	27.270/0.730	24.902/0.606	25.103/0.706
x 4	Bicubic	15.349/0.473	14.148/0.386	19.164/0.463	20.491/0.601	18.412/0.218	17.631/0.496
	SRCNN	17.880/0.560	15.854/0.465	20.822/0.574	21.096/0.653	19.031/0.294	18.937/0.589
	VDSR	18.237/0.558	16.672/0.493	21.197/0.587	21.639/0.648	19.753/0.342	19.426/0.569
	DRRN	18.549/0.567	16.872/0.506	21.468/0.601	21.948/0.653	20.088/0.387	19.708/0.581
	RCAN	18.934/0.561	17.394/0.514	21.897/0.604	22.681/0.671	20.473/0.406	20.187/0.604
	CDC	19.619/0.572	17.918/0.539	22.161/0.618	23.017/0.669	20.943/0.417	20.884/0.647
	CVA	19.706/0.577	17.975/0.541	22.187/0.617	23.064/0.673	21.084/0.416	20.803/0.565
	TTSR	19.743/0.584	18.039/0.548	22.203/0.619	23.137/0.679	21.103/0.420	20.845/0.570
	MICU	19.802/0.590	18.107/0.550	22.326/0.622	23.207/0.681	21.132/0.423	20.915/0.573
	Ours	19.809/0.595	18.129/0.558	22.351/0.620	23.250/0.686	21.191/0.422	20.946/0.656

#### 4.2. Subjective visual analysis

To observe more intuitively the performance of the proposed traffic sign image super-resolution reconstruction algorithm with other algorithms in human subjective vision, the test images of traffic signs in five different environments used in the Experiment 4.1 section are visualized for comparative analysis. Since the PSNR and SSIM metrics after image reconstruction of different algorithms have been recorded in Section 4.1, only the visual effect graphs of the reconstructed images are shown in this section and analyzed comparatively. Figures 7–11 show the comparison maps of the super-resolution reconstruction effect of the traffic sign test set images in five environments, namely, early morning, midday with strong illumination, nighttime under lights, rainy day, and foggy day, respectively, after 4-fold downsampling. Figure 7 shows the reconstruction effect of each algorithm in the morning when there is no sunlight irradiation, and this paper's algorithm reconstruction of the

traffic sign text is closest to the high-resolution image, slightly clearer than the original HR image, compared with other algorithms. Figure 8 shows the visualization of the reconstruction effect of the traffic sign at noon with strong light reflection, and this paper’s algorithm and the recently released CVA, TTSR, and MICU algorithms have better reconstruction results, and compared to this paper’s image these are softer, indicating that there is more detailed information. Figure 9 shows the reconstruction effect of different algorithms for traffic signs under the influence of multiple light sources at night, and it can be seen that the CDC, CVA, TTSR, and MICU algorithms can identify the content of “Jiaoda East Road”, while for other algorithms the text of the traffic signs is fuzzier and more difficult to recognize, and thus this paper’s algorithm has the highest clarity and strongest visibility. Figure 10 shows the foggy weather traffic signs and present white mask fuzzy state. It tests this paper’s algorithm and other algorithms in the important traffic signs pattern reconstruction effect, and the test results show that, it prohibits the sounding of horns and prohibit overtaking and other signs in the "15" characters. This paper’s algorithm in the reconstruction of the edges of the more clearly, and there are no jagged edges. Figure 11 is a comparison of the reconstruction effect of the fuzzy traffic sign image under a high-speed driving environment in the rain, and this paper's algorithm is closest to the original HR image in terms of color, brightness, and clarity, and thus it can be easier to identify the traffic sign on the signage prompt information.



**Figure 7.** Comparison of the 4-fold reconstruction rate between this algorithm and other algorithms in a morning light environment.

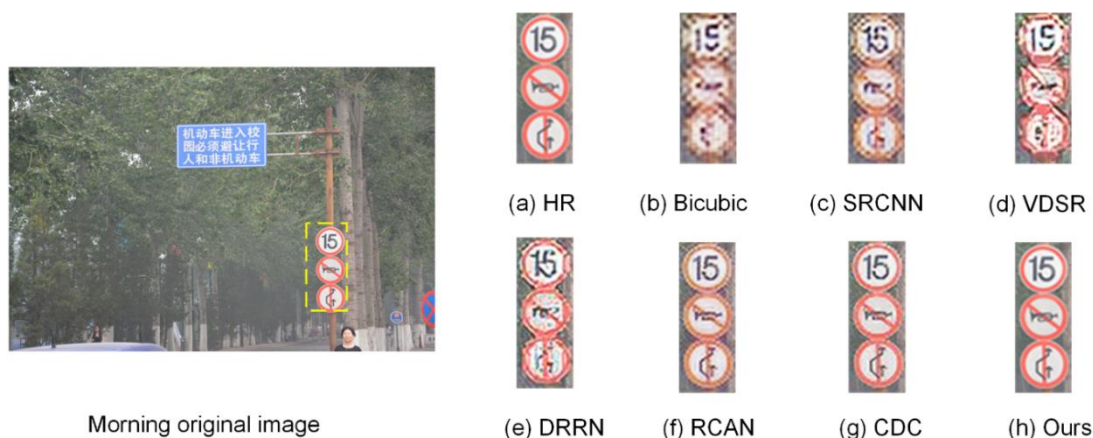


**Figure 8.** Comparison of the effect of the 4-fold reconstruction rate between this algorithm and other algorithms under a midday strong illumination environment.

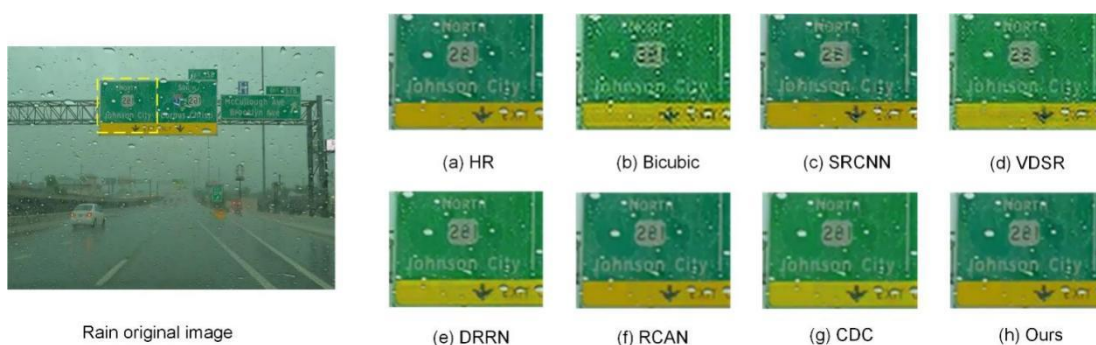




**Figure 9.** Comparison of the effect of this paper’s algorithm and other algorithms’ 4-fold reconstruction rate at night under a poor line-of-sight environment.



**Figure 10.** Comparison of the effect of the 4-fold reconstruction rate between this algorithm and other algorithms in a foggy visibility reduction environment.



**Figure 11.** Comparison of the effect of this algorithm and other algorithms’ 4-fold reconstruction rate on rainy days and bad weather environments.

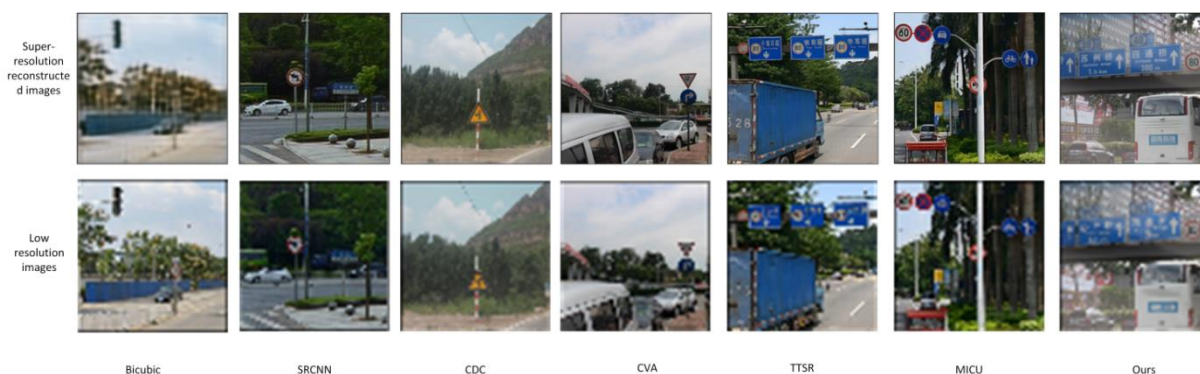
To compare the quality of deep learning algorithms in the reconstruction of high-frequency information such as texture, details, edges, and other high-frequency information of feature maps of deep networks, calling the `torchvision.utils.save_image` method in PyTorch allows observing the image of feature layers in any layer of the network to study the network’s ability to learn the features and the

effect of the representation of the network. In this paper, we use this method in the last layer of the model reconstruction layer. The latest MICU algorithm with better performance is selected to compare the feature map details with the 2-fold reconstruction rate of this paper's algorithm to observe the reconstruction effect of the deep learning model on high-frequency detail information, and the comparison results are shown in Figure 12. Observation of Figure 12 finds that the high-frequency detail information features of this paper's algorithm are more obvious than the MICU algorithm; the details, textures, and edges are clearer and the pixels are more coherent, which indicates that this paper's algorithm has a better reconstruction effect in the detailed texture part.



**Figure 12.** Comparison of the details of the feature maps of the last up-sampled layer of the reconstructed layer.

To further compare the effect of different reconstruction algorithms on random image reconstruction in natural scenes, the images in the test set are arbitrarily selected for 4-fold double-three times interpolation downsampling to obtain a low-resolution image, and then the low-resolution image is reconstructed into a super-resolution image using the traditional method, the early deep learning method, and the recently published reconstruction method. The reconstructed image is compared with the original high-resolution image to observe the reconstruction effect, as shown in Figure 13.



**Figure 13.** Effect of different algorithms in reconstructing 4-fold downsampled low-resolution images in randomized natural scenes.

In Figure 13, the first row of images is reconstructed by different algorithms, the second row of

low-resolution images simulated by the three times interpolation downsampling method, the second row of images is overall very fuzzy where it is difficult to recognize the traffic signs and text in the picture. Observing the first row of reconstructed images by different algorithms, it can be seen that the traditional Bicubic algorithm reconstructs the image with very poor effect, and the legibility is even lower than that of the four times downsampled images. The SRCNN and CDC methods are earlier deep learning methods, and the reconstruction effect is significantly improved, but there is still an obvious fuzzy visual sensation, and the detailed reconstruction effect needs to be further improved. The CVA, TTSR, and MICU models are the latest image super-resolution reconstruction algorithms based on the deep learning method, and their image reconstruction quality is significantly better than the earlier SRCNN and CDC methods in terms of clarity, light and dark coordination, detailed texture, and other aspects, but the text recognition clarity still needs to further improvement. Finally, for the method proposed in this paper, it can be seen that for foggy days, the reconstruction effect is still better, the text and traffic signs are clearer. Compared with the latest proposed reconstruction algorithms, this method has a better practical visual reconstruction effect.

From the comparative results of these experiments, the algorithm in this paper outperforms the earlier mainstream representative algorithms and the latest released algorithms in terms of objective indexes and subjective visual effects in the task of super-resolution reconstruction of dimly lit blurred traffic sign images. The fusion of multi-layer network features effectively integrates the high-frequency information with the high-level semantic information, which provides the reconstruction layer with rich detailed texture edge features and semantic recognition features. The dual inverse channel sub-pixel convolutional up-sampling strategy provides the reconstructed image with richer pixel and spatial information, which can make full use of the different levels of network fusion to reconstruct the detailed edges and subtle texture structures.

### 4.3. Ablation experiments

To demonstrate the strong contribution of the Attention Residual module (AR), the Important Feature Fusion Strategy (IFFS) at the heterogeneous network layer, and the Dual Inverse Channel Subpixel Convolutional Algorithm (DICSC) to the overall network model reconstruction performance, the following experiments provide a point-by-point comparison of the three improvement components by the control variable method. In this experiment, different model PSNR and SSIM values are computed using the difficult 4-fold super-resolution for the traffic sign test set images, and the image super-pixel reconstruction network model combined with six convolutional feature extraction modules and sub-pixel convolutional upsampling modules is used as the baseline model; model a indicates that the attentional residuals module is used without the feature fusion strategy and the dual inverse channel sub-pixel convolution method; model b indicates that the feature fusion strategy without using the attention residual module and the dual inverse channel subpixel convolution method; model c indicates that the dual inverse channel subpixel convolution up-sampling method is used without using the attention residual module and the feature fusion strategy; model d (Ours) is the final model proposed in this paper, which contains the attention residual module, the feature fusion strategy, and the dual inverse channel subpixel convolution up-sampling method, and the results of the ablation experiments are shown in Table 3.

**Table 3.** Effect of different modules on model performance.

Module			Model	Morning	Noon	Night	Rain	Fog	Mean
AR	IFFS	DICSC		x4	x4	x4	x4	x4	x4
				PSNR/SSIM	PSNR/SSIM	PSNR/SSIM	PSNR/SSIM	PSNR/SSIM	PSNR/SSIM
×	×	×	Standard	17.654/0.439	16.957/0.483	21.049/0.538	21.791/0.583	19.860/0.399	18.884/0.532
√	×	×	a	18.568/0.532	17.323/0.520	21.639/0.599	22.664/0.631	20.291/0.407	19.591/0.607
×	√	×	b	18.554/0.539	17.309/0.518	21.597/0.583	22.667/0.636	20.285/0.401	19.579/0.586
×	×	√	c	18.574/0.547	17.316/0.529	21.623/0.591	22.672/0.643	20.294/0.408	19.582/0.599
√	√	√	Ours	<b>19.809/0.595</b>	<b>18.129/0.558</b>	<b>22.351/0.620</b>	<b>23.250/0.686</b>	<b>21.191/0.422</b>	<b>20.946/0.656</b>

Observing and analyzing the experimental results in Table 3, to verify the importance of the attention residual module in the process of feature extraction, the residual connection avoids the problem of gradient vanishing in the process of backpropagation, and at the same time ensures that the high-frequency information features are effectively retained. The attention mechanism strengthens the expression of the important features and reduces the amount of the low-frequency redundant information transmission to optimize the feature extraction ability, which can be supported by the significant improvement of the evaluation indexes in the model as compared to the baseline models, and the PSNR and SSIM evaluation indexes can also be corroborated. Similarly, observing that the evaluation indexes of model b and model c are both improved compared to the baseline model indicates that the feature fusion strategy in model b and the dual inverse channel sub-pixel convolutional up-sampling method in model c both play an active role in enriching the reconstruction layer with detailed high-frequency and high-semantic information, enhancing the reconstruction performance, and the fusion of important features in different feature layers. The dual inverse channel sub-pixel convolution method enriches the detailed edge pixels and ensures multiple spatial information in the unit expansion region, which ultimately improves the visibility of the local physical properties of the image, such as the details and textures. Finally, when all the modules work together, the PSNR and SSIM reach the highest level of 20.946dB and 0.656, respectively, which is a better performance than using any of the modules singularly. This shows that each module has its advantages and also promote each other to make the overall model reach the best state.

## 5. Statement of conclusions and limitations

In this paper, a super-resolution reconstruction algorithm for dim fuzzy traffic sign images is proposed, which incorporates three main functional modules, namely, an attention residual network, important feature fusion module, and double inverse channel sub-pixel convolution structure. The attention residual network is designed for the main feature extraction, which adopts the residual bottleneck structure to deepen the network to improve the nonlinear expression ability and fit more complex features while avoiding the gradient disappearance that leads to the model failure, and the attention mechanism strengthens the expression of important high-frequency information and suppresses the propagation of redundant information. In addition, the important hierarchical feature fusion module adequately fuses the important features of different feature layers, and the high-frequency and high-semantic features are combined with the important features of different layers. The high-frequency features and high-semantic features are fully fused to enrich the reconstructed feature layer and improve the expression of detailed texture features. Finally, the dual inverse channel sub-pixel convolution algorithm is used for up-sampling to realize super-resolution reconstruction of the image, which makes the spatial layout of pixels in the reconstructed image more reasonable, attenuates the obvious pixel misalignment phenomenon, and makes the image visibility stronger. The experimental evaluation of the proposed algorithm model was carried out on the Chinese traffic sign image dataset, and the experiments show that this paper's algorithm achieves better performance on the traffic sign image dataset which is more difficult to recognize. Compared with other algorithms, it performs better in both objective evaluation indexes and subjective visualization, and this paper's algorithm has a better reconstruction effect in the edge details, and it has a better practical visualization effect. The method in this paper improves the accuracy of gray and fuzzy traffic sign recognition in complex and harsh environments, which improves the reliability of safe driving aids and enriches human-computer interaction technology in automatic driving.

The research in this paper has some limitations. The experimental subjects in this paper do not consider a wider range of geographical traffic signage, as well as the effects of more seasons, climates, periods, inclement weather, and other environments on the quality of super-resolution reconstruction of traffic sign images, and the model's generalization ability needs to be further verified. In addition, this paper did not select all the reconstruction algorithms released in the last three years for performance comparison and only selected three of the newer released models for comparative experiments, so it is not possible to determine the performance of all the image super-resolution reconstruction algorithms in this field. In subsequent work, a wider range of traffic sign image training data will be collected and other reconstruction algorithms will be verified for their performance on the task of super-resolution reconstruction of traffic sign images.

### Use of AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

### Acknowledgments

Thank you very much for the two major funding sources provided by Xinyu Hu, namely the

National Natural Science Foundation of China (No. 61976083); And Hubei Province Key R&D Program of China (No. 2022BBA0016).

### Conflict of interest

The authors declare there is no conflict of interest.

### References

1. K. Zhou, Y. Zhan, D. Fu, Learning region-based attention network for traffic sign recognition, *Sensors*, **21** (2021), 686. <https://doi.org/10.3390/s21030686>
2. Z. Liu, Y. Cai, H. Wang, L. Chen, H. Gao, Y. Jia, et al., Robust target recognition and tracking of self-driving cars with radar and camera information fusion under severe weather conditions. *IEEE T. Intell. Transp. Syst.*, **23** (2021), 6640–6653. <https://doi.org/10.1109/TITS.2021.3059674>
3. M. Hnewa, H. Radha, Object detection under rainy conditions for autonomous vehicles: A review of state-of-the-art and emerging techniques, *IEEE Signal Proc. Mag.*, **38** (2020), 53–67. <https://doi.org/10.1109/MSP.2020.2984801>
4. O. Soufi, Z. Aarab, F. Belouadha, Benchmark of deep learning models for single image super-resolution (SISR), In: *2022 2nd International Conference on Innovative Research in Applied Science, Engineering and Technology (IRASET)*, 2022. <https://doi.org/10.1109/IRASET52964.2022.9738274>
5. K. Li, S. Yang, R. Dong, X. Wang, J. Huang, Survey of single image super-resolution reconstruction, *IET Image Processing*, **14** (2022), 2273–2290. <https://doi.org/10.1049/iet-ipr.2019.1438>
6. D. Qiu, Y. Cheng, X. Wang, Medical image super-resolution reconstruction algorithms based on deep learning: A survey, *Comput. Meth. Prog. Bio.*, **238** (2023), 107590. <https://doi.org/10.1016/j.cmpb.2023.107590>
7. L. Zhang, R. Dong, S. Yuan, W. Li, J. Zheng, H. Fu, Making low-resolution satellite images reborn: A deep learning approach for super-resolution building extraction, *Remote Sens.*, **13** (2021), 2872. <https://doi.org/10.3390/rs13152872>
8. H. Chen, X. He, L. Qing, Y. Wu, C. Ren, R. E. Sheriff, et al., Real-world single image super-resolution: A brief review, *Inform. Fusion*, **79** (2022), 124–145. <https://doi.org/10.1016/j.inffus.2021.09.005>
9. S. C. Park, M. K. Park, M. G. Kang, Super-resolution image reconstruction: A technical overview, *IEEE Signal Proc. Mag.*, **20** (2003), 21–36. <https://doi.org/10.1109/MSP.2003.1203207>
10. D. O. Bager, J. Leuschner, M. Schmidt, Computed tomography reconstruction using deep image prior and learned reconstruction methods, *Inverse Probl.*, **36** (2020), 094004. <https://doi.org/10.1088/1361-6420/aba415>
11. J. Xiao, H. Yong, L. Zhang, Degradation model learning for real-world single image super-resolution, In: *Computer Vision–ACCV 2020*, 2020. [https://doi.org/10.1007/978-3-030-69532-3\\_6](https://doi.org/10.1007/978-3-030-69532-3_6)
12. P. Wu, J. Liu, M. Li, Y. Sun, F. Shen, Fast sparse coding networks for anomaly detection in videos, *Pattern Recogn.*, **107** (2020), 107515. <https://doi.org/10.1016/j.patcog.2020.107515>

13. J. Li, S. Wei, W. Dai, Combination of manifold learning and deep learning algorithms for mid-term electrical load forecasting, *IEEE T. Neur. Net. Lear. Syst.*, **34** (2023), 2584–2593. <https://doi.org/10.1109/TNNLS.2021.3106968>
14. F. Deeba, S. Kun, F. Ali Dharejo, Y. Zhou, Sparse representation based computed tomography images reconstruction by coupled dictionary learning algorithm, *IET Image Process.*, **14** (2020), 2365–2375. <https://doi.org/10.1049/iet-ipr.2019.1312>
15. C. Dong, C. C. Loy, K. He, X. Tang, Learning a deep convolutional network for image super-resolution, In: *Computer Vision–ECCV 2014*, 2014, 184–199. [https://doi.org/10.1007/978-3-319-10593-2\\_13](https://doi.org/10.1007/978-3-319-10593-2_13)
16. J. Kim, J. K. Lee, K. M. Lee, Deeply-recursive convolutional network for image super-resolution, In: *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, 1637–1645. <https://doi.org/10.1109/CVPR.2016.181>
17. Y. Tai, J. Yang, X. Liu, Image super-resolution via deep recursive residual network, In: *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, 2790–2798. <https://doi.org/10.1109/CVPR.2017.298>
18. Y. Zhang, K. Li, K. Li, L. Wang, B. Zhong, Y. Fu, Image super-resolution using very deep residual channel attention networks, In: *Computer Vision–ECCV 2018*, 2018, 294–310. [https://doi.org/10.1007/978-3-030-01234-2\\_18](https://doi.org/10.1007/978-3-030-01234-2_18)
19. T. Dai, J. Cai, Y. Zhang, S. T. Xia, L. Zhang, Second-order attention network for single image super-resolution, In: *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019, 11057–11066. <https://doi.org/10.1109/cvpr.2019.01132>
20. P. Wei, Z. Xie, H. Lu, Z. Zhan, Q. Ye, W. Zuo, et al., Component divide-and-conquer for real-world image super-resolution, In: *Computer Vision–ECCV 2020*, 2020, 101–117. [https://doi.org/10.1007/978-3-030-58598-3\\_7](https://doi.org/10.1007/978-3-030-58598-3_7)
21. W. Zhang, W. Zhao, J. Li, P. Zhuang, H. Sun, Y. Xu, et al., CVANet: Cascaded visual attention network for single image super-resolution, *Neural Networks*, **170** (2024), 622–634. <https://doi.org/10.1016/j.neunet.2023.11.049>
22. Y. Wang, S. Jin, Z. Yang, H. Guan, Y. Ren, K. Cheng, et al., TTSR: A transformer-based topography neural network for digital elevation model super-resolution, *IEEE T. Geosci. Remote Sens.*, **62** (2024), 4403179. <https://doi.org/10.1109/TGRS.2024.3360489>
23. Y. Chen, R. Xia, K. Yang, K. Zou, MICU: Image super-resolution via multi-level information compensation and U-net, *Expert Syst. Appl.*, **245** (2024), 123111. <https://doi.org/10.1016/j.eswa.2023.123111>
24. Z. H. Qu, Y. M. Shao, T. M. Deng, J. Zhu, X. H. Song, Traffic sign detection and recognition under complex lighting conditions, *Laser. Optoelectron. P.*, **56** (2019), 231009. <https://doi.org/10.3788/LOP56.231009>
25. X. G. Zhang, X. L. Liu, J. Li, H. D. Wang, Real-time detection and recognition of speed limit traffic signs under BP neural network, *J. Xidian Univ.*, **45** (2018), 136–142. <https://doi.org/10.3969/j.issn.1001-2400.2018.05.022>
26. G. Z. Xu, Y. Zhou, B. Dong, C. C. Liao, Traffic signage recognition based on improved cascade R-CNN. *Sens. Microsyst.*, **40** (2021), 142–145+153. [https://doi.org/10.13873/j.1000-9787\(2021\)05-0142-04](https://doi.org/10.13873/j.1000-9787(2021)05-0142-04)



27. L. Liu, S. Lu, R. Zhong, B. Wu, Y. Yao, Q. Zhang, et al., Computing systems for autonomous driving: State of the art and challenges, *IEEE Internet Things J.*, **8** (2021), 6469–6486. <https://doi.org/10.1109/JIOT.2020.3043716>
28. H. Singh, A. Kathuria, Analyzing driver behavior under naturalistic driving conditions: A review. *Accident Anal. Prev.*, **150** (2021), 105908. <https://doi.org/10.1016/j.aap.2020.105908>
29. S. Woo, J. Park, J. Y. Lee, I. S. Kweon, CBAM: Convolutional block attention module, In: *Computer Vision—ECCV 2018*, 2018, 3–19. [https://doi.org/10.1007/978-3-030-01234-2\\_1](https://doi.org/10.1007/978-3-030-01234-2_1)
30. J. Deng, W. Dong, R. Socher, L. J. Li, K. Li, F. F. Li, Imagenet: A large-scale hierarchical image database, In: *2009 IEEE Conference on Computer Vision and Pattern Recognition*, 2009, 248–255. <https://doi.org/10.1109/CVPR.2009.5206848>
31. X. Wang, K. Yu, C. Dong, C. C. Loy, Recovering realistic texture in image super-resolution by deep spatial feature transform, In: *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018, 606–615. <https://doi.org/10.1109/CVPR.2018.00070>
32. A. Ignatov, N. Kobyshev, R. Timofte, K. Vanhoey, DSLR-quality photos on mobile devices with deep convolutional networks, In: *2017 IEEE International Conference on Computer Vision (ICCV)*, 2017, 3297–3305. <https://doi.org/10.1109/ICCV.2017.355>
33. J. Hu, L. Shen, S. Albanie, G. Sun, E. Wu, Squeeze-and-excitation networks, *IEEE T. Pattern. Anal.*, **42** (2020), 2011–2023. <https://doi.org/10.1109/tpami.2019.2913372>
34. Z. Cui, N. Wang, Y. Su, W. Zhang, Y. Lan, A. Li, ECANet: Enhanced context aggregation network for single image dehazing, *Signal Image Video P.*, **17** (2023), 471–479. <https://doi.org/10.1007/s11760-022-02252-w>
35. J. Xu, Z. Li, B. Du, M. Zhang, J. Liu, Reluplex made more practical: Leaky ReLU, In: *2020 IEEE Symposium on Computers and Communications (ISCC)*, 2020, 1–7. <https://doi.org/10.1109/ISCC50000.2020.9219587>
36. F. Nie, H. Huang, X. Cai, C. Ding, Efficient and robust feature selection via joint  $\ell_{2,1}$ -norms minimization, *Adv. Neural Inform. Processing Syst.*, 2010.
37. A. Hore, D. Ziou, Image quality metrics: PSNR vs. SSIM, In: *2010 20th International Conference on Pattern Recognition*, 2010, 2366–2369. <https://doi.org/10.1109/ICPR.2010.579>
38. D. Han, Comparison of commonly used image interpolation methods, In: *Proceedings of the 2nd International Conference on Computer Science and Electronics Engineering (ICCSEE 2013)*, 2013, 1556–1559. <https://doi.org/10.2991/iccsee.2013.391>



AIMS Press

© 2024 the Author(s), licensee AIMS Press. This is an open access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>)