Neural Networks 170 (2024) 622–634

Contents lists available at ScienceDirect

Neural Networks

journal homepage: www.elsevier.com/locate/neunet

Full length article

CVANet: Cascaded visual attention network for single image super-resolution

Weidong Zhang [a], Wenyi Zhao [b,*], Jia Li [a], Peixian Zhuang [c], Haihan Sun [d], Yibo Xu [b], Chongyi Li [e]

[a] *School of Information Engineering, Henan Institute of Science and Technology, Xinxiang, 453003, China*
[b] *School of Artificial Intelligence, Beijing University of Posts and Telecommunications, Beijing, 100876, China*
[c] *School of Automation and Electrical Engineering, University of Science and Technology Beijing, 100084, China*
[d] *School of Engineering, University of Tasmania, Tasmania, 7005, Australia*
[e] *School of Computer Science, Nankai University, Tianjing, 300073, China*

ARTICLE INFO

Keywords:
Super-resolution
Feature attention
Channel attention
Pixel attention
Closely-related modules

ABSTRACT

Deep convolutional neural networks (DCNNs) have exhibited excellent feature extraction and detail reconstruction capabilities for single image super-resolution (SISR). Nevertheless, most previous DCNN-based methods do not fully utilize the complementary strengths between feature maps, channels, and pixels. Therefore, it hinders the ability of DCNNs to represent abundant features. To tackle the aforementioned issues, we present a Cascaded Visual Attention Network for SISR called CVANet, which simulates the visual attention mechanism of the human eyes to focus on the reconstruction process of details. Specifically, we first designed a trainable feature attention module (FAM) for feature-level attention learning. Afterward, we introduce a channel attention module (CAM) to reinforce feature maps under channel-level attention learning. Meanwhile, we propose a pixel attention module (PAM) that adaptively selects representative features from the previous layers, which are utilized to generate a high-resolution image. Satisfactory, our CVANet can effectively improve the resolution of images by exploring the feature representation capabilities of different modules and the visual perception properties of the human eyes. Extensive experiments with different methods on four benchmarks demonstrate that our CVANet outperforms the state-of-the-art (SOTA) methods in subjective visual perception, PSNR, and SSIM.The code will be made available https://github.com/WilyZhao8/CVANet.

1. Introduction

Single image super-resolution (SISR) is an interesting image reconstruction in computer vision, which has attracted significant attention from researchers. Usually, SISR aims to produce a clear and high-resolution (HR) image from the corresponding low-resolution (LR) image. Notably, SISR is a hot research area due to its widespread application in industrial applications (Qin, Chen, Jeon, & Yang, 2023), satellite video (Chen, Zhang, & Huang, 2022), data control systems (Chandrasekar, Radhika, & Zhu, 2022b; Tamil Thendral, Ganesh Babu, Chandrasekar, & Cao, 2022), state stability (Chandrasekar, Radhika, & Zhu, 2022a; Radhika, Chandrasekar, Vijayakumar, & Zhu, 2023), etc. Unfortunately, SISR is an ill-posed issue because there exists various solutions for converting LR images to HR images. To deal with the inverse issue, various learning-based methods have been extensively applied to explore the non-linear mapping relationship from LR images to HR images.

Recently, numerous techniques built on deep convolutional neural networks have demonstrated high-quality reconstruction capabilities for SISR (Kim, Lee, & Lee, 2016a; Shi et al., 2016), which makes the generated HR images with richer texture details. Dong, Loy, and Tang (2016) first presented a super-resolution convolutional neural network (SRCNN) with a three-layer CNN structure to learn the mapping from LR images to HR images. Afterward, some researchers designed more deep super-resolution networks inspired by residual learning (Lan et al., 2021; Li, Fang, Mei, & Zhang, 2018), recursive learning (Kim, Lee, & Lee, 2016b; Tai, Yang, & Liu, 2017), ensemble learning (Jiang et al., 2020), and weakly supervised learning (Ren, Wang, & Zhang, 2023a; Zhang, Deng, et al., 2023), which achieved better reconfiguration performance than the SRCNN. Unfortunately, these methods consider modifying the depth and width of the network, thus resulting in an additional computational cost of SR reconstruction. Influenced by this attention mechanism, Zhang, Li, et al. (2018) fused the SENet (Hu, Shen, Albanie, Sun, & Wu, 2020) to design a channel-wise attention network for extracting channel attention features. However, the global average pooling employed in channel attention may lose some features. To alleviate this issue, Dai, Cai, Zhang, Xia, and Zhang (2019) introduced a novel second-order attention network to replace channel attention. Additionally, Tian et al. (2021) presented a coarse-to-fine

[*] Corresponding author.
E-mail address: zwy@bupt.edu.cn (W. Zhao).

https://doi.org/10.1016/j.neunet.2023.11.049
Received 9 March 2023; Received in revised form 27 September 2023; Accepted 22 November 2023
Available online 24 November 2023

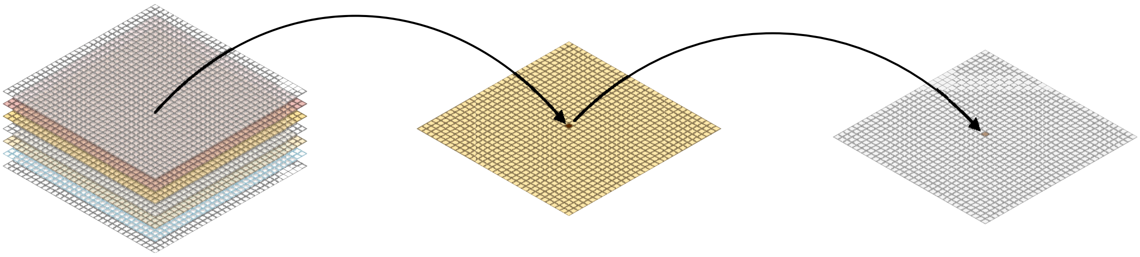


Fig. 1. How to exploit the most valuable pixel in several feature maps. CVANet first focuses on the subjective vision of the human eyes in terms of multiple fields of view, called **feature attention**. Next, which focuses on the specific channel in the field of view, called **channel attention**. Finally, which focuses on the spatial location of the most valuable pixel in the channel, called **pixel attention**.

CNN to aggregate complementary texture information and overcame the instability of the training model. Generally speaking, the SISR models with attention mechanisms demonstrate superior reconfiguration capabilities for super-resolution compared to traditional SISR models.

Although the SISR models mentioned above can achieve better reconfiguration performance, they still suffer from some limitations compared to human visual perception: (1) Most of the existing SISR modes cannot obtain larger receptive fields for a single convolution layer. Hence, relationships between features that are far apart are not easily established. For instance, MSRN (Li et al., 2018) uses a multi-scale strategy to increase the receptive field, but reconstruction results are unsatisfactory because it needs to consider multi-scale features sufficiently. (2) Most of the existing SISR modes usually utilize skip-connection to the end layer for each module in the middle layer, which fully uses deep and shallow features. Unfortunately, the usefulness of these features for the reconstructed results has yet to be fully considered. Therefore, the low-level contour and high-level semantic features cannot be fully utilized. (3) Existing methods use the Pixel-Shuffle (Caballero et al., 2017) strategy, which is challenging to select finer feature maps. To sum up, most existing methods lead to a loss of information at the pixel level, and these limitations will hinder the feature representation capability of DCNNs.

To address the aforementioned issues, we present a Cascaded Visual Attention Network for SISR, called CVANet. Concretely, three closely-related modules optimized for SISR tasks are presented. First, a feature attention module is introduced to obtain larger perceptual fields and multi-scale feature maps while consuming fewer computational resources, which utilizes feature maps of different scales to learn the interrelationships between feature maps adaptively. Next, a channel attention module with global average pooling is proposed to combine skip-connection to fully exploit deep and shallow features, which ensures that the feature maps processed by it have strong channel attention features allowing our CVANet to focus on more helpful feature maps and improve the model reconstruction capability. Finally, we propose a pixel attention module to provide pixel attention feature mappings for each layer and adaptively select valuable feature mappings to generate super-resolution images. Besides, Fig. 1 demonstrates the hierarchical progression of these three attentional modules. In short, the outstanding contributions of our work are highlighted as follows.

- We present a cascaded visual attention super-resolution network designed to emulate human visual perceptual characteristics to refine the reconstruction processes, which effectively addresses the issue of underutilization of features in existing methods. Extensive experiments with different methods on four benchmarks demonstrate that our CVANet has better reconstruction performance.
- We have developed three collaborative attention modules to obtain refined reconstructed features. Firstly, the FAM captures a broad perceptual field. Secondly, the CAM directs the model's attention toward pivotal features within the channels. Lastly, the pixel attention module adaptively selects significant pixels for the reconstructed super-resolution images.
- We introduce three closely interlinked modules using a cascading strategy for image reconstruction tasks, which enables our CVANet to acquire a more resilient feature representation. Additionally, our proposed modules exhibit strong versatility and can be seamlessly integrated into other tasks to enhance the overall performance of the CVANet.

2. Related work

Currently, various SISR models have been extensively presented. In this section, we focus on DCNN-based and Attention-based SISR models that are relevant to our model.

2.1. DCNN-based SISR models

Recently, DCNN-based methods have achieved state-of-the-art reconstruction performance for SISIR than traditional methods (Ahn, Kang, & Sohn, 2018; Wang et al., 2021; Zhang, Deng, et al., 2023). Unfortunately, most existing methods focus only on the width or depth of the network, and they without fully exploit the relationship between the feature maps, channels, and pixels. For example, Zhang, Tian, et al. (2018) designed a residual dense network (RDN) to reconstruct a high super-resolution image, which employed residual dense block to capture local features. However, the RDN ignores the fine features between channels and pixels. To extract fine features, Li et al. (2018) designed a multi-scale residual network (MSRN) to alleviate this issue. Nevertheless, the features extracted by MSRN depending on multiple calculations of scale rates and only focus on the attention features of the feature level. He et al. (2019) designed an ordinary differential equation-inspired network (OISR), which thanks to the two modules LF-block and RK-block. Dai et al. (2019) attempted to replace the global average pooling with second-order statistics to exploit the relationship between channels, but it ignores the reconstruction of texture details. To alleviate the issue, Tian et al. (2021) gathered complementary contextual information to construct a coarse-to-fine CNN to overcome the instability of training. Fang, Li, and Zeng (2020) designed a soft-edge assisted network to further extract edge textures for detail enhancement of high-resolution images. Sun et al. (2021) designed a weighted multi-scale residual network to better tradeoff between experimental results and calculated costs, which adaptively uses the feature representations at different scale spaces by dilated convolutions with different scales. Wang, Su, et al. (2022) presented a plug-in reparameterized dynamic distillation network to improve the reconfiguration performance and the computational cost trade-off. Esmaeilzehi, Ahmad, and Swamy (2022) proposed an ultralight-weight convolutional neural network, which designed a three-prior formulation of the optimization issue for SISR. Sun, Pan, and Tang (2022) employed the large depth-wise convolution to reconstruct fine details. However, most methods are limited by the effective learning ability of the network, which cannot deal with the relationship between feature level, channel level, and pixel level attention characteristics.

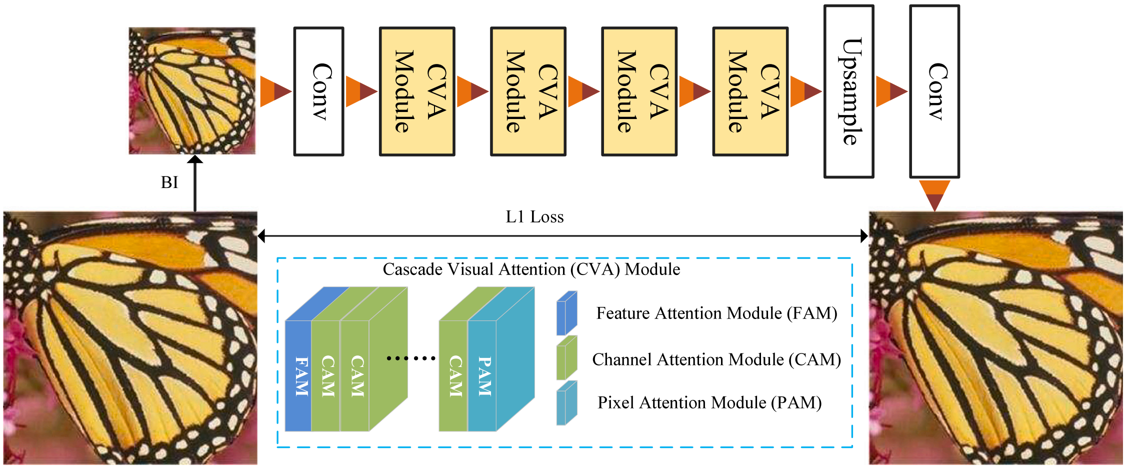

Conv
CVA Module
CVA Module
CVA Module
CVA Module
Upsample
Conv
BI
L1 Loss
Cascade Visual Attention (CVA) Module
FAM
CAM
CAM
CAM
PAM
Feature Attention Module (FAM)
Channel Attention Module (CAM)
Pixel Attention Module (PAM)

Fig. 2. The architecture of our proposed Cascaded Visual Attention Network (CVANet), which consists of three stages: feature attention, channel attention, and pixel attention. The three closely related modules use a cascading strategy for image reconstruction tasks, allowing our CVANet to learn a more robust feature representation. Besides, our presented modules have good applicability and can be easily embedded into other tasks to improve the performance of the model.

2.2. Attention-based SISR models

Noteworthy, the visual attention is a unique human brain signal processing mechanism. Precisely, human vision first observes the global image and obtain the target area that needs attention. Meanwhile, more attention is devoted to this area to get more detailed target characteristics and suppress other useless features. Although deep and wide networks capture image feature information through tedious learning strategies (Cao, Chandrasekar, Radhika, & Vijayakumar, 2023; Kim et al., 2016a; Lai, Huang, Ahuja, & Yang, 2017; Rakkiyappan, Chandrasekar, & Cao, 2014), it is challenging to focus on valuable information (Ren, Kong, Zhang, & Wang, 2023; Ren, Wang, & Zhang, 2023b). Recently, attention strategy has been gradually applied to DCNN to improve the superior performance of some visual tasks, such as image classification (Zhang, Li, et al., 2022), object detection (Li et al., 2021), image enhancement (Zhang, Zhuang, et al., 2022; Zhuang, Wu, Porikli, & Li, 2022). Gao et al. (2021) proposed a multi-scale backbone architecture for many computer vision tasks, the specialized design benefits from the fact that it forces CNN to pay attention to effective representation capabilities in each layer. Hu et al. (2020) proposed SENet for SISR, which forced CNN to pay more attention to the relationship between channels, and it achieved signification performance gains for image classification and other computer vision tasks. Li, Wang, Hu, and Yang (2020) proposed SKNet, which selected the most informative feature maps for subsequent network processing. Wu et al. (2021) proposed a multi-grained attention network, which fully explores the advantages of attention mechanisms and multi-scale, while the multi-grained attention blocks generate multi-scale features at each layer. Yan et al. (2021) designed a novel graph attention network to fully exploit the internal patch recurrence in a high-resolution image, which focuses on refining low-level representation abilities with high-level information. Song and Zhong (2022) presented a lightweight local–global attention network, which employs the self-attention mechanism to dependencies for each pixel. The local–global attention module to refine the local and global features. Fang, Lin, Chen, and Zeng (2022) designed a hybrid network of CNN and transformer for SISR, which introduces spatial attention to the entire network to improve the performance of high-resolution image reconstruction. However, these attention mechanisms only focus on a certain level of features. Therefore, a more comprehensive method that combines several attention mechanisms is proposed to simulate the human visual characteristics to refine the reconstruction in SISR.

3. Methodology

In our work, we design a Cascaded Visual Attention Network (CVANet) for SISR. From Fig. 2, CVANet can be divided into three stages: feature attention, channel attention, and pixel attention. In the first stage, we employ the feature attention module to acquire a large perceptual field. In the second stage, the channel attention module focuses on representative features in the channel (Zhao, Yang, Pan, & Li, 2021; Zhao et al., 2023). In the final stage, the pixel attention module adaptively selects representative pixels for a reconstructed SR image. We designed the three closely-related modules using a cascading strategy for image reconstruction tasks, which allows the proposed CVANet to learn a more robust feature representation.

3.1. Network framework

CVANet expects to predict a HR image I_{SR} from a LR image I_{LR}. Let us denote I_{LR} and I_{SR} as the low-resolution and high-resolution images of our proposed CVANet, and I_{HR} as the ground-truth. Following previous works (Han, Zheng, Chen, & Wang, 2022; Lei et al., 2021), we only employ a single convolutional layer to capture the shallow features from the low-resolution image, which is expressed as:

$$F_0 = H_{\mathrm{SF}}\left(I_{\mathrm{LR}}\right), \tag{1}$$

where $H_{\mathrm{SF}}(\cdot)$ represents the shallow feature extraction convolution process, F_0 denotes the shallow features, which are utilized for following shallow feature extraction and cascaded visual attention learning. Subsequently, the Eq. (2) is redefined as:

$$F_{\mathrm{CVA}} = H_{\mathrm{CVA}}\left(F_0\right), \tag{2}$$

where H_{CVA} denotes our proposed cascaded visual attention structure, including a feature attention module (FAM), a channel attention module (CAM), and a pixel attention module (PAM). F_{CVA} represents the feature maps extracted via H_{CVA} operation. Thanks to our flexible design, our proposed cascaded visual attention structure can be efficiently embedded into other network structures for different vision tasks. Luo, Li, Urtasun, and Zemel (2016) have demonstrated that there was a massive gap among the theoretical receptive field and the effective receptive field in deep convolutional neural networks, which would get a more robust and effective model if it were adjusted in the shallow layers. Consequently, we employ the FAM to adjust the receptive field

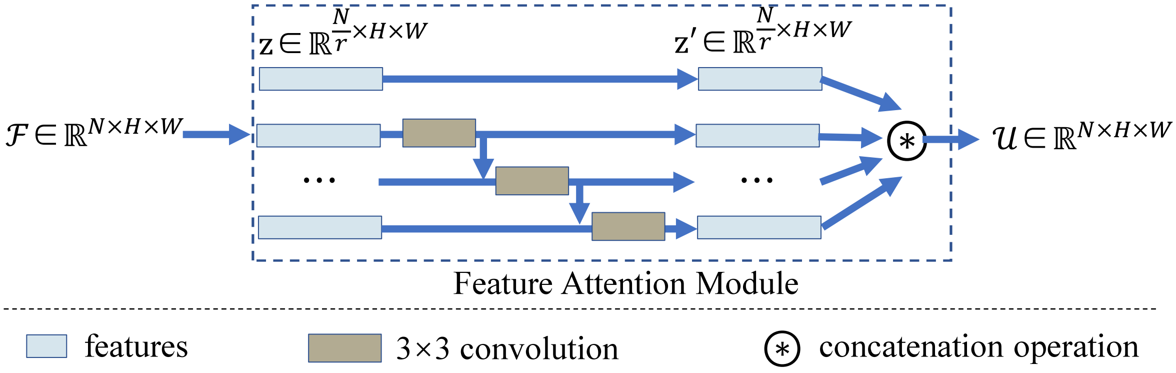

$z \in \mathbb{R}^{\frac{N}{r} \times H \times W}$
$z' \in \mathbb{R}^{\frac{N}{r} \times H \times W}$
$\mathcal{F} \in \mathbb{R}^{N \times H \times W}$
$\mathcal{U} \in \mathbb{R}^{N \times H \times W}$
Feature Attention Module
features
3×3 convolution
concatenation operation

Fig. 3. The network architecture of FAM, which consists of hierarchically cascaded convolutional modules. Initially, the feature maps are uniformly divided into several groups, with one group of convolutional kernels dedicated to extracting features from the preceding group. Subsequently, these features are passed to the next group for further processing, and all the features are concatenated together before the final output to restore the dimensionality of the input features.

gap in $H_{\text{SF}}(\cdot)$. Afterwards, the F_{CVA} upscaled via an upscale module as:

$$F_{\text{UP}} = H_{\text{UP}}\left(F_{\text{CVA}}\right), \tag{3}$$

where $H_{\text{SF}}(\cdot)$ and F_{UP} represent the upscale module and the upscaled features, respectively. At last, the upscaled features are reconstructed by one convolutional layer, which is expressed as:

$$I_{\text{SR}} = H_{\text{REC}}\left(F_{\text{UP}}\right) = H_{\text{CVA}}\left(I_{\text{LR}}\right), \tag{4}$$

where $H_{\text{REC}}(\cdot)$ and $H_{\text{CVA}}(\cdot)$ represent the reconstruction layer and the cascaded visual attention structure function, respectively. For the upscale module, there are some choices, such as the deconvolution layer (Dumoulin, Shlens, & Kudlur, 2017), nearest-neighbor up-sampling with convolutional layers (Wang, Su, et al., 2022), and efficient sub-pixel convolutional neural network (ESPCN) (Caballero et al., 2017). In our work, we utilize a post-upscaling strategy, and the upscaling method has been demonstrated to be more efficient than pre-upscaling SR methods (Zhang, Li, et al., 2018).

3.2. Feature attention module

We give more details about our designed FAM in Fig. 3. Our FAM includes multi-scale feature map attention and local residual learning to form a continuous memory mechanism. Moreover, we remove batch-normalization (BN) in the proposed FAM to adapt the super-resolution tasks and save computing resources. Multi-scale feature attention is designed by changing the various receptive fields at a single granular level.

In the field of super-resolution, each convolution layer is impossible to obtain a larger receptive field in existing methods. However, Luo et al. (2016) have proved that it leads to a more robust and efficient model if we can obtain a large receptive field. Therefore, it is necessary to construct a model with large receptive field in shallow layer to extract informative feature maps. Some methods (Li et al., 2018; Li, Lin, Dong, & Zhang, 2020) combine multi-scale strategies, which can increase the receptive field to a certain extent. However, because some methods underutilize multi-scale information, they contain only feature-level attentional features to the extent that the results may be more satisfactory. Inspired by Tian et al. (2022), we first obtain the shallow feature maps by a 1 × 1 convolution operation, and then the input feature maps are divided into several groups. For the first group, it is directly equal to the corresponding output by skip-connection. For other groups, a group of filters is used to extract features from the previous group and then sent to the next group for subsequent processing, and an illustration is presented in Fig. 3. If all the groups are processed, feature maps from all the groups and skip-connection are concatenated and sent to a 1 × 1 convolution. It has been demonstrated that the structure has a powerful feature extraction ability, which can enhance the multi-scale representation ability (Gao et al., 2021). Moreover, the structure also provides multi-scale feature and a large receptive field to the subsequent networks. We also utilize a long skip connection in the FAM since it can alleviate the model degradation (Kingma & Ba, 2015; Luo et al., 2016).

3.3. Channel attention module

Existing CNN-based SR methods rarely consider the interdependencies between channels. RCAN (Zhang, Li, et al., 2018) introduced SENet (Hu et al., 2020) to refine the channel-wise features for SISR. Nevertheless, RCAN (Zhang, Li, et al., 2018) only exploits channel attention features by introducing global average pooling, which will lose important information as only channel dimension characteristics are considered. To alleviate the issue, we combine feature level and pixel level statistics in our model to exploit the relationship between channels.

Specifically, let us denote $\mathbf{F} = \left[\mathbf{f}_1, \dots, \mathbf{f}_c\right]$ as a $H \times W \times C$ features with C channels and size of $H \times W$, and the channel-wise statistics are formulated as:

$$z_c = H_{\text{GAP}}(\mathbf{y}_c) = \frac{1}{C}\sum_{i}^{C}\mathbf{y}_c(i), \tag{5}$$

where z_c is the cth dimension of $z \in R^{C\times 1}$, z can be obtained by shrinking $\hat{\mathbf{Y}}=[\mathbf{y}_1, \cdots, \mathbf{y}_C]$, and $H_{\text{GAP}}(\cdot)$ denotes the global average pooling. From Fig. 4, the reduction and increase rates of the channel downscaling module and the channel upscaling module after global pooling are $r\left(W_D\right)$ and $r\left(W_U\right)$, respectively. W_D and W_U consist of convolution layers, and the f is a sigmoid function to satisfy two conditions that the gate mechanism discussed in SENet (Hu et al., 2020). Meanwhile, the function must have the ability to learn non-linear interactions among channels, which must learn a non-mutually-exclusive relationship. Therefore, the result of channel attention module with global average pooling as:

$$\hat{X} = X \bullet \left\{f(W_U\delta(W_D z))\right\}, \tag{6}$$

where X and $\hat{X}$ is the input and output of CAM, $\delta(\cdot)$ denotes ReLU function and z represent the output of $H_{\text{GAP}}(\cdot)$. Inspired by Kim et al. (2016b), we set $r = 16$ in the CAM. Meanwhile, the CAM-processed features have channel statistical attention features that are similar to the human visual system at the channel level.

3.4. Pixel attention module

Most existing CNN-based SR methods embed upscaling modules in the last few layers to obtain a better tradeoff between computational burden and performance. However, all of them treat the features extracted by the previous layers equally, which makes the model unable

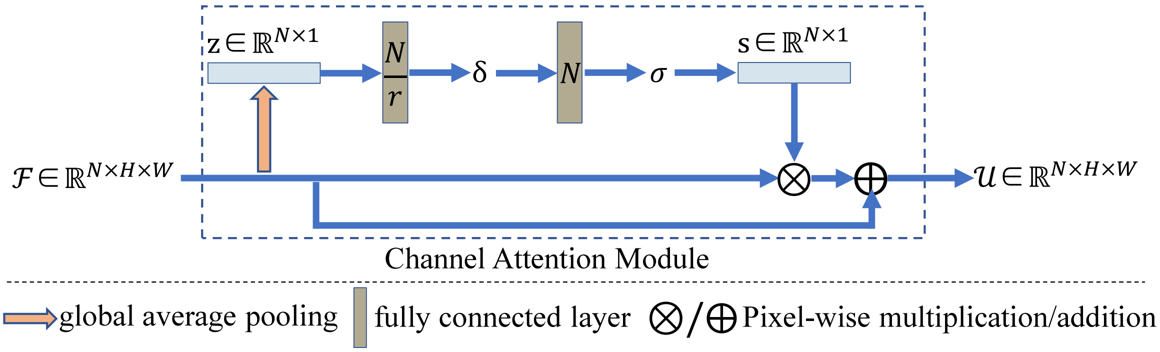


Fig. 4. The network architecture of CAM involves first subjecting the input features to global average pooling, followed by dimensionality reduction and expansion processes to extract crucial information from the features. Subsequently, the features, activated through the ReLU function, are multiplied with the original features, and the obtained results are added back to the original features to yield the output features. This module significantly enhances the model's ability to extract features along the channel dimension.

to exploit the feature interdependencies and cannot fully use the previous feature maps. Since the last few layers are crucial to make up the final pixels in SR images, choosing which feature map to form the final pixels is essential and necessary. That is, pixels are the factors that make up the final SR image. Choosing pixels is to select the feature maps that will generate the pixels. Therefore, it is essential to reconstruct SR images using the pixel attention mechanism, which selects feature maps to reconstruct SR images.

How to generate different feature maps for selection is a crucial step. There exists two concerns: on the one hand, the features to be selected must contain different information, so the chosen features are more robust and more representative. On the other hand, the selection mechanism must be dynamic, allowing each kernel to adaptively adjust its selection via the representative of the input information.

Based on these analyses, we introduce a sparse self-attention mechanism to implement our proposed PAM. As shown in Fig. 5, support $\mathbf{X} = [x_1, \dots, x_c, \dots x_C]$ be an input, which has C feature map with size of $H \times W$, $\mathbf{Y}$ is the output of the pixel attention module (PAM). We first split $\mathbf{X}$ into two branches, the upper branch splits the input features into buckets by hashing, and the bottom branch calculates self-attention for every query within each bucket. For further efficiency, we introduce sparse self-attention instead of traditional full-range self-attention. Specifically, given a query location i, the corresponding output of full-range self-attention y_i can be expressed as:

$$y_i = \sum_{j=1}^{hw} \frac{f\left(x_i, x_j\right)}{\sum_{\hat{j}=1}^{hw} f\left(x_i, x_{\hat{j}}\right)} g\left(x_j\right), \tag{7}$$

where x_i, x_j, and $x_{\hat{j}}$ donates the pixel-wise features at location i, j, and $\hat{j}$ on $\mathbf{X}$. $f()$ calculate the mutual similarities and $g()$ is a feature transformation operation. The Eq. (7) can be also formulated as $y_i = D\alpha_i$, where $D = \left[g\left(x_1\right), \dots, g\left(x_{hw}\right)\right] \in R^{c \times hw}$ and $\alpha_i = \left[f\left(x_i, x_1\right), \dots, f\left(x_i, x_{\mathrm{hw}}\right)\right] \in R^{\mathrm{hw}}$.

To construct a high-efficiency model, the calculation can be reduced by reducing the number of non-zero terms of α up to k. Thus, the sparse self-attention is derived as:

$$\begin{aligned} y_i &= D\alpha_i \quad \text{s.t.} \ \left\|\alpha_i\right\|_0 \le k \\ &= \sum_{j \in \delta_i} \frac{f\left(x_i, x_j\right)}{\sum_{j \in \delta_i} f\left(x_i, x_j\right)} g\left(x_j\right), \end{aligned} \tag{8}$$

where δ_i is non-zero terms of α_i i.e., $\delta_i = \left\{j \mid \alpha_i[j] \neq 0\right\}$, which constrains the identified locations where sparse self-attention can be computed from. $\alpha_i[j]$ denotes the jth element in α_i. Following image super-resolution with non-local sparse attention (Chen et al., 2021), δ_i can be defined as $\delta_i = \left\{j \mid \alpha_i[j] \neq 0\right\}$, where $h(x) = \arg\max_i(\hat{x}) = \arg\max_i \left(\mathbf{A}\left(\frac{x}{\|x\|_2}\right)\right)$ and $\mathbf{A}$ represents a rotation matrix. It is obvious that the sparse self-attention module can save considerable computational overhead. Since the pixels in SR images are obtained by feature maps after PAM, the final results contain more comprehensive characteristics. PAM is proposed to simulate the human being's visual system at the pixel level. Based on these attention modules and long-skip-connection structure, we construct a new SISR model named CVANet to demonstrate the effectiveness of our CVANet.

Algorithm 1: Reconstruct an SR image (I_{SR}) from a LR image (I_{LR}).

```
Input: $\mathbf{I}_{\mathrm{LR}}$
while Training do
    Shallow Features Extractor: $\mathbf{F}_0 = H_{\mathrm{SF}}(\mathbf{I}_{\mathrm{LR}})$;
    Do $H_{\mathrm{CVA}}$ : Cascaded visual attention steps;
        FAM: $\mathbf{F}_{\mathrm{FAM}} = H_{\mathrm{FAM}}(\mathbf{F_0})$;
        CAM: $\mathbf{F}_{\mathrm{CAM}} = H_{\mathrm{FAM}}(\mathbf{F}_{\mathrm{FAM}})$;
        PAM: $\mathbf{F}_{\mathrm{PAM}} = H_{\mathrm{FAM}}(\mathbf{F}_{\mathrm{FCM}})$;
    Reconstructed SR image: $\mathbf{I}_{\mathrm{SR}^*} = \mathrm{Conv}(\mathrm{Up}(\mathrm{Conv}(\mathbf{F}^*)))$;
    Update the loss function: $L(\Theta)$;
end while
$\mathbf{I}_{\mathrm{LR}} = \mathbf{I}_{\mathrm{LR}^*}$;
Output: $\mathbf{I}_{\mathrm{LR}}$;
```

3.5. Loss function

A loss function is used to accumulate similarities and differences between I_{HR} and I_{SR}, and guide the model to optimize the parameters. Recently, some loss functions have gradually been applied to SR, such as pixel loss (Ahn et al., 2018; Lai et al., 2017), adversarial loss (Ran et al., 2023; Zhang, Liu, Dong, Zhang, & Yuan, 2020), cycle consistency loss (Gao et al., 2021; Kim et al., 2022) and total variation loss (Li, Wang, Hu, & Yang, 2019; Luo et al., 2016). The L_1 loss is the minimum absolute deviation to sum the fundamental difference between the actual and target values. In the supervised image super-resolution task, the goal is to make the generated image as close as possible to the real high-resolution image. Therefore, the L_1 loss is used to calculate the error of the value of SR and HR corresponding to the pixel position. In this paper, we use L_1 loss, which is belongs to pixel loss. Let us denote $\left\{I^i_{LR}, I^i_{HR}\right\}^N_{i=1}$ as a training set, which includes N pairs of LR and HR, the aim of training CVANet is to minimize the L_1 loss, which is expressed as:

$$L(\Theta) = \frac{1}{N}\sum_{i=1}^{N} \| H_{SA}(I^i_{LR}) - I^i_{HR}\|_1, \tag{9}$$

where Θ denotes the parameters that our CVANet needs to learn. Additionally, The detail of our proposed CVANet for SISR in Algorithm 1. Specifically, H is operations, while F^* is output feature maps. Conv and Up are convolution and up-scaling operations, respectively.

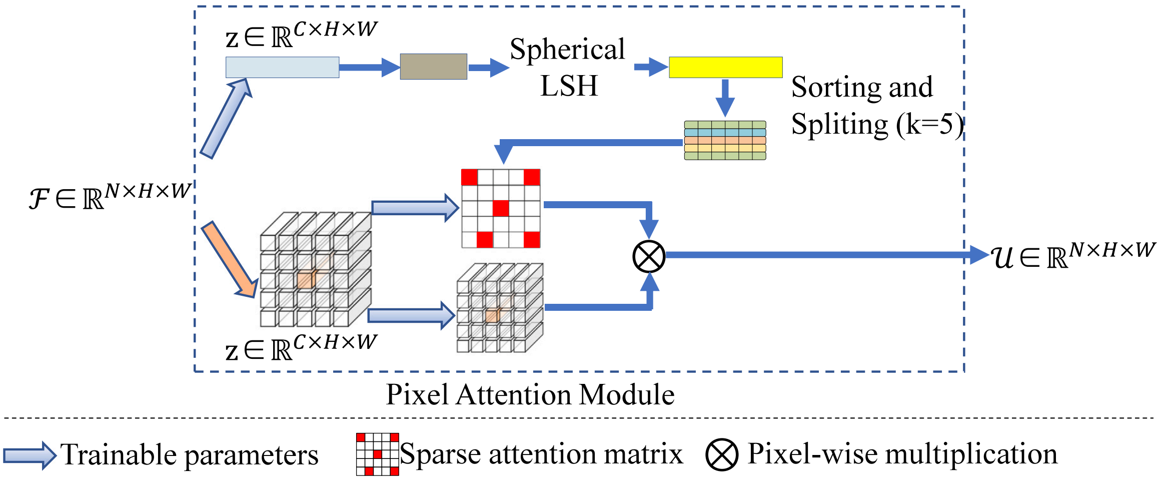


Fig. 5. The network architecture of the PAM, which adaptively selects representative features from the previous layers. The ⇨ and ⇨ indicate identity mapping and trainable parameters, while ➞ represents the direction of feature flow. The upper branch of this module splits the input features into buckets by hashing, and the bottom branch calculates self-attention for every query within each bucket. Through this sparse self-attention instead of traditional full-range self-attention, PAM can save considerable computational overhead.

4. Experimental results and analysis

In this section, we comprehensively evaluate our designed CVANet on several benchmarks. Specifically, we begin with an overview of the datasets and evaluation metrics, followed by implementation details. Whereafter, our CVANet and several SOTA methods were compared qualitatively and quantitatively. Finally, we conduct the ablation analysis to demonstrate the effect of our designed modules and extend our model to other frameworks to verify its scalability.

4.1. Datasets and evaluation metrics

In our experiment, we utilize 800 high-resolution training images selected from the DIV2K dataset (Agustsson & Timofte, 2017) as the training dataset. We crop high-resolution patches and downsample via the bicubic operation to simulate the corresponding low-resolution patches. In testing, we adopt standard benchmarks of Set5 (Bevilacqua, Roumy, Guillemot, & Morel, 2012), Set14 (Zeyde, Elad, & Protter, 2012), BSD100 (Martin, Fowlkes, Tal, & Malik, 2001), and Urban100 (Huang, Singh, & Ahuja, 2015). Additionally, we employ two commonly-utilized image quality metrics PSNR and SSIM to evaluate SR performance. PSNR is utilized to measure the similarity between the initial image and the reconstructed image. Mathematically, it can be expressed as:

$$\mathrm{PSNR} = 10 \cdot \log_{10}{\left(\mathrm{MAX}_{\mathbf{I}}^{2}/\mathrm{MSE}\right)}, \tag{10}$$

where $\mathrm{MSE} = \frac{1}{mn}\sum_{i=o}^{m-1}\sum_{j=0}^{n-1}[\mathbf{I}(i,j) - \mathbf{K}(i,j)]^2$, $\mathrm{MAX}_{\mathbf{I}}^{2}$ is the maximum possible pixel value of the image, m and n are the width and height of the image. The higher value of PSNR suggests a better quality of the reconstructed image.

SSIM is employed to measure the structural similarity between the initial image and the reconstructed image. Mathematically, it can be expressed as:

$$\mathrm{SSIM}(O,T) = \frac{\left(2\mu_O\mu_T + c_1\right)\left(2\sigma_{OT} + c_2\right)}{\left(\mu_O^2 + \mu_T^2 + c_1\right)\left(\sigma_O^2 + \sigma_T^2 + c_2\right)}, \tag{11}$$

where μ_O and μ_T are the mean values of the original and target images, σ_O and σ_T are the variances of the original and target images, and σ_{OT} are covariances the of the original and target images. The higher value of SSIM means that the reconstructed image is closer to the initial image. In general, the higher the index of SSIM and PSNR, the better the quality of reconstructed images in engineering applications.

4.2. Implementation details

We implement the proposed CVANet via the PyTorch with two RTX3090 GPU. During training, we perform random rotation operations of 90°, 180°, 270°, and horizontally flipping on the training images to augment the dataset. In each training batch, 16 LR color patches with the resolution of 48 × 48 are provided as inputs. Our CVANet is optimized by ADAM (Kingma & Ba, 2015) with the parameters $\beta_1 = 0.9$, $\beta_2 = 0.999$, and $\varepsilon = 10^{-8}$. The initial learning rate is set to 10^{-4} and then decreases to half every 200 epochs, and the total epoch is 1000. We chose MSRN (Li et al., 2018) and RCAN (Zhang, Li, et al., 2018) as our baselines to prove the scalability of our designed modules, and the proposed three attention modules are embedded in MSRN (Li et al., 2018) and RCAN (Zhang, Li, et al., 2018). Since we presented the FAM to optimize the receptive field characteristics, when using MSRN (Li et al., 2018) as the baseline, the FAM, CAM, and PAM are all embedded in the model. The model has a larger receptive field when using RCAN (Zhang, Li, et al., 2018) as the benchmark model because RCAN (Zhang, Li, et al., 2018) contains a deeper network structure. Therefore, only CAM and PAM are embedded in the model to balance the calculation and performance.

4.3. Comparisons with SOTA methods

To comprehensively verify the performance of our proposed CVANet, our method is compared with 20 SISR methods both quantitatively and qualitatively, including deep network methods: DRCN (Kim et al., 2016b), LapSRN (Lai et al., 2017), VDSR (Kim et al., 2016a), DRRN (Tai et al., 2017), MSRN (Li et al., 2018), OISR (He et al., 2019), SeaNet (Fang et al., 2020), TPCNN (Esmaeilzehi et al., 2022), and WDRN (Xin et al., 2022); lightweight methods: CARN (Ahn et al., 2018), IDN (Hui et al., 2018), WMRN (Sun et al., 2021), DDistill (Wang, Su, et al., 2022), and ShuffleMixer (Sun et al., 2022); attention methods: MGAN (Wu et al., 2021), SRGAT (Yan et al., 2021), HNCT (Fang et al., 2022), SMSR (Wang et al., 2021), and LGAN (Song & Zhong, 2022). Meanwhile, the authors provide the source code and running parameters of all comparison methods to obtain the best results.

(1) **Quantitative Evaluation:** Table 1 reports the perception comparison results of our CVANet and other SOTA methods on several benchmark datasets with scale factor ×2, ×3, and ×4. Meanwhile, the results of SOTA methods are derived from their papers. Nevertheless, there are also some results derived from the test output of their papers providing source code. As reported in quantitative comparisons, our proposed CVANet with similar and in general highest

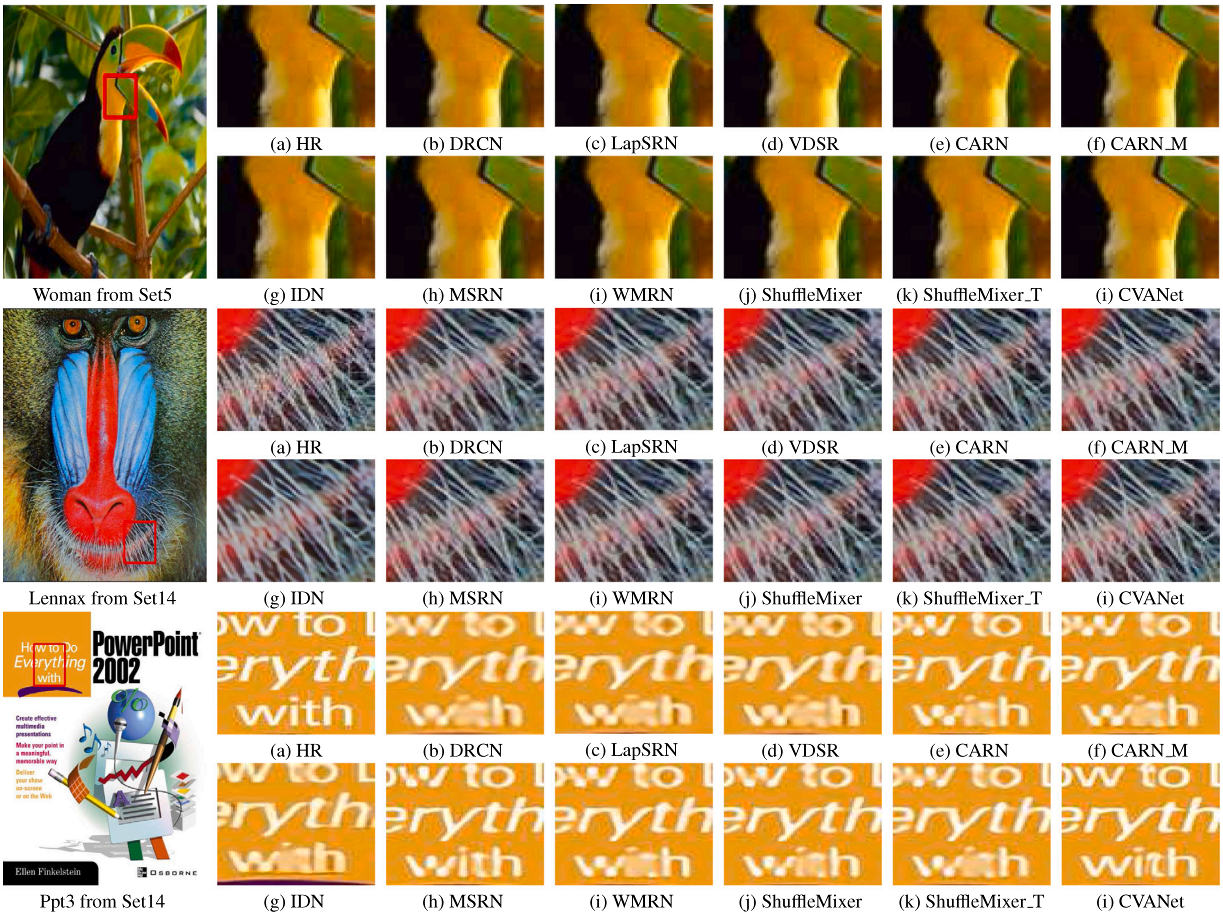
Table 1
Quantitative evaluation of the average values of PSNR and SSIM of SOTA methods on several benchmarks with scale factor ×2, ×3, and ×4. Red and blue suggest optimal and suboptimal results.

Method	Scale	Pub. & Year	Set5		Set14		B100		Urban100	
			PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
BICUBIC (Dengwen, 2010)	×2	–	33.66	0.9299	30.24	0.8688	29.56	0.8431	26.88	0.8403
DRCN (Kim et al., 2016b)		CVPR 2016	37.63	0.9588	33.04	0.9188	31.90	0.8942	30.75	0.9133
LapSRN (Lai et al., 2017)		CVPR 2017	37.52	0.9591	32.99	0.9124	31.80	0.8952	30.41	0.9103
VDSR (Kim et al., 2016a)		CVPR 2016	37.53	0.9587	33.03	0.9124	31.90	0.8960	30.76	0.9140
IDN (Hui, Wang, & Gao, 2018)		CVPR 2018	37.83	0.9600	33.30	0.9148	32.08	0.8985	31.27	0.9196
CARN (Ahn et al., 2018)		ECCV 2018	37.76	0.9590	33.52	0.9166	32.09	0.8978	31.92	0.9256
MSRN (Li et al., 2018)		ECCV 2018	38.08	0.9605	33.74	0.9170	32.23	0.9013	32.22	0.9326
SeaNet (Fang et al., 2020)		TIP 2020	38.15	0.9611	33.86	0.9198	32.31	0.9013	32.68	0.9332
SMSR (Wang et al., 2021)		CVPR 2021	38.00	0.9601	33.64	0.9179	32.17	0.8990	32.19	0.9284
WMRN (Sun et al., 2021)		IEEE JAS 2021	37.83	0.9599	33.41	0.9162	32.08	0.8984	31.68	0.9241
MGAN (Wu et al., 2021)		TMM 2021	38.21	0.9614	33.91	0.9205	32.33	0.9015	32.95	0.9354
DDistill (Wang, Su, et al., 2022)		TMM 2022	38.08	0.9608	33.73	0.9195	32.25	0.9007	32.39	0.9301
SRGAT (Yan et al., 2021)		TIP 2021	38.20	0.9610	33.93	0.9201	32.34	0.9014	32.90	0.9359
ShuffleMixer (Sun et al., 2022)		NeurIPS 2022	38.01	0.9606	33.63	0.9180	32.17	0.8995	31.89	0.9257
HNCT (Fang et al., 2022)		CVPR 2022	38.08	09608	33.65	0.9182	32.22	0.9001	32.22	0.9294
LGAN (Song & Zhong, 2022)		ACCV 2022	38.13	0.9612	33.95	0.9221	32.32	0.9017	32.81	0.9343
TPCNN (Esmaeilzehi et al., 2022)		TAI 2022	38.03	0.9613	33.67	0.9187	32.25	0.9014	31.76	0.9257
WDRN (Xin et al., 2022)		TNNLS 2022	38.19	0.9631	33.39	0.9212	32.27	0.9014	32.64	0.9372
CVANet			38.30	0.9616	34.17	0.9224	32.41	0.9024	33.37	0.9388
BICUBIC (Dengwen, 2010)	×3	–	30.39	0.8682	27.55	0.7742	27.21	0.7385	24.46	0.7349
DRCN (Kim et al., 2016b)		CVPR 2016	33.82	0.9226	29.76	0.8311	28.80	0.7963	27.15	0.8276
LapSRN (Lai et al., 2017)		CVPR 2017	33.81	0.9220	29.79	0.8325	28.82	0.7980	27.07	0.8275
VDSR (Kim et al., 2016a)		CVPR 2016	33.66	0.9213	29.77	0.8314	28.82	0.7976	27.14	0.8279
IDN (Hui et al., 2018)		CVPR 2018	34.11	0.9253	29.99	0.8354	28.95	0.8013	27.42	0.8359
CARN (Ahn et al., 2018)		ECCV 2018	34.29	0.9255	30.29	0.8407	29.06	0.8034	28.06	0.8493
MSRN (Li et al., 2018)		ECCV 2018	34.38	0.9262	30.34	0.8395	29.08	0.8041	28.08	0.8554
SeaNet (Fang et al., 2020)		TIP 2020	34.65	0.9290	30.53	0.8461	29.23	0.8081	28.68	0.8620
SMSR (Wang et al., 2021)		CVPR 2021	34.40	0.9270	30.33	0.8412	29.10	0.8050	28.25	0.8536
WMRN (Sun et al., 2021)		IEEE JAS 2021	34.11	0.9251	30.17	0.8390	28.98	0.8021	27.80	0.8448
MGAN (Wu et al., 2021)		TMM 2021	34.75	0.9299	30.60	0.8474	29.29	0.8098	28.82	0.8651
DDistill (Wang, Su, et al., 2022)		TMM 2022	34.43	0.9276	30.39	0.8432	29.16	0.8070	28.31	0.8546
SRGAT (Yan et al., 2021)		TIP 2021	34.75	0.9297	30.63	0.8474	29.29	0.8099	28.95	0.8666
ShuffleMixer (Sun et al., 2022)		NeurIPS 2022	34.40	0.9272	30.37	0.8423	29.12	0.8051	28.08	0.8498
HNCT (Fang et al., 2022)		CVPR 2022	34.47	0.9275	30.44	0.8439	29.15	0.8067	28.28	0.8557
LGAN (Song & Zhong, 2022)		ACCV 2022	34.62	0.9286	30.60	0.8463	29.24	0.8092	28.79	0.8646
TPCNN (Esmaeilzehi et al., 2022)		TAI 2022	34.43	0.9281	30.48	0.8451	29.16	0.8082	28.03	0.8514
WDRN (Xin et al., 2022)		TNNLS 2022	34.62	0.9292	30.50	0.8454	29.20	0.8085	28.59	0.8625
CVANet			34.84	0.9303	30.70	0.8489	29.33	0.8110	29.21	0.8717
BICUBIC (Dengwen, 2010)	×4	–	28.42	0.8104	26.00	0.7027	25.96	0.6675	23.14	0.6577
DRCN (Kim et al., 2016b)		CVPR 2016	31.53	0.8854	28.02	0.7670	27.32	0.7233	25.14	0.7510
LapSRN (Lai et al., 2017)		CVPR 2017	31.54	0.8852	28.09	0.7700	27.32	0.7275	25.21	0.7562
VDSR (Kim et al., 2016a)		CVPR 2016	31.35	0.8838	28.01	0.7674	27.29	0.7251	25.18	0.7524
IDN (Hui et al., 2018)		CVPR 2018	31.82	0.8903	28.25	0.7730	27.41	0.7297	25.41	0.7632
CARN (Ahn et al., 2018)		ECCV 2018	32.13	0.8937	28.60	0.7806	27.58	0.7349	26.07	0.7837
MSRN (Li et al., 2018)		ECCV 2018	32.07	0.8903	28.60	0.7751	27.52	0.7273	26.04	0.7896
SeaNet (Fang et al., 2020)		TIP 2020	32.44	0.8981	28.81	0.7872	27.70	0.7399	26.50	0.7976
SMSR (Wang et al., 2021)		CVPR 2021	32.12	0.8932	28.55	0.7808	27.55	0.7351	26.11	0.7868
WMRN (Sun et al., 2021)		IEEE JAS 2021	32.00	0.8925	28.47	0.7786	27.49	0.7328	25.89	0.7789
MGAN (Wu et al., 2021)		TMM 2021	32.57	0.8993	28.85	0.7874	27.75	0.7415	26.68	0.8027
DDistill (Wang, Su, et al., 2022)		TMM 2022	32.29	0.8961	28.69	0.7833	27.65	0.7385	26.26	0.7893
SRGAT (Yan et al., 2021)		TIP 2021	32.57	0.8997	28.86	0.7879	26.76	0.7421	26.76	0.8052
ShuffleMixer (Sun et al., 2022)		NeurIPS 2022	32.21	0.8953	28.66	0.7827	27.61	0.7366	26.08	0.7835
HNCT (Fang et al., 2022)		CVPR 2022	32.31	0.8957	28.71	0.7834	27.63	0.7381	26.20	0.7896
LGAN (Song & Zhong, 2022)		ACCV 2022	32.48	0.8984	28.83	0.7864	27.71	0.7416	26.63	0.8022
TPCNN (Esmaeilzehi et al., 2022)		TAI 2022	32.14	0.8957	28.72	0.7846	27.62	0.7381	26.00	0.7835
WDRN (Xin et al., 2022)		TNNLS 2022	32.43	0.8985	28.75	0.7862	27.65	0.7384	26.41	0.7975
CVANet			32.59	0.9001	28.92	0.7896	27.77	0.7437	26.96	0.8116

PSNR and SSIM on four benchmarks with scale factor ×2, ×3, and ×4, which outperforms almost all compared methods. In additional, our proposed CVANet can embedded in different baselines to improve the performance.

(2) **Qualitative Evaluation:** We quantitatively evaluated the scores of our CVANet on the challenge images from four benchmarks with scale factor ×4 in Figs. 6–8. From these results, we observe that most contrast methods produce blurred edges. In contrast, our CVANet can reconstruct more natural and clearer textures than SOTA methods. These satisfactory results also demonstrate that our CVANet has superior performance for image reconstruction by simulating the cascaded attention mechanism of human visual perception.

(3) **Discussions:** Difference to Res2Net (Gao et al., 2021), SENet (Hu et al., 2020), and non-local attention module (Chen et al., 2021). Res2Net (Gao et al., 2021) introduces a efficient and flexible multi-scale convolutional neural network, and this module extends the receptive fields at a more granular level. There are some differences between Res2Net (Gao et al., 2021) and our proposed FAM. First, Res2Net (Gao et al., 2021) extends the receptive fields at a more granular level for a wide variety of high-level visual tasks (Lei et al., 2023; Wang, Ma, Jiang, & Zhang, 2022). In contrast, our method incorporates Res2Net (Gao et al., 2021) operation in deep CNN for image SR and other low-level computer vision tasks. Second, Res2Net (Gao et al., 2021) only considers the receptive fields, which means it only

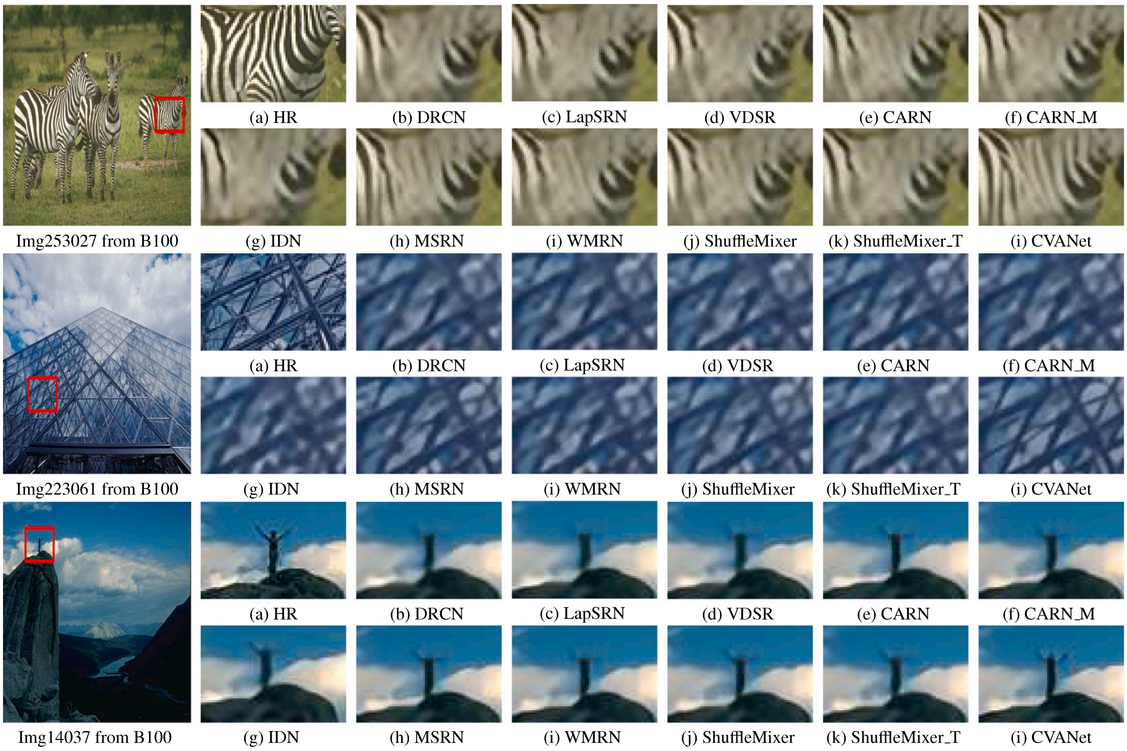


Woman from Set5
(a) HR (b) DRCN (c) LapSRN (d) VDSR (e) CARN (f) CARN_M
(g) IDN (h) MSRN (i) WMRN (j) ShuffleMixer (k) ShuffleMixer_T (i) CVANet

Lennax from Set14
(a) HR (b) DRCN (c) LapSRN (d) VDSR (e) CARN (f) CARN_M
(g) IDN (h) MSRN (i) WMRN (j) ShuffleMixer (k) ShuffleMixer_T (i) CVANet

Ppt3 from Set14
(a) HR (b) DRCN (c) LapSRN (d) VDSR (e) CARN (f) CARN_M
(g) IDN (h) MSRN (i) WMRN (j) ShuffleMixer (k) ShuffleMixer_T (i) CVANet

Fig. 6. Visual perception comparison results of different SR methods on the Set5 and Set14 datasets with ×4 SR images.



Img253027 from B100
(a) HR (b) DRCN (c) LapSRN (d) VDSR (e) CARN (f) CARN_M
(g) IDN (h) MSRN (i) WMRN (j) ShuffleMixer (k) ShuffleMixer_T (i) CVANet

Img223061 from B100
(a) HR (b) DRCN (c) LapSRN (d) VDSR (e) CARN (f) CARN_M
(g) IDN (h) MSRN (i) WMRN (j) ShuffleMixer (k) ShuffleMixer_T (i) CVANet

Img14037 from B100
(a) HR (b) DRCN (c) LapSRN (d) VDSR (e) CARN (f) CARN_M
(g) IDN (h) MSRN (i) WMRN (j) ShuffleMixer (k) ShuffleMixer_T (i) CVANet

Fig. 7. Visual perception comparison results of different SR methods on the B100 dataset with ×4 SR images.

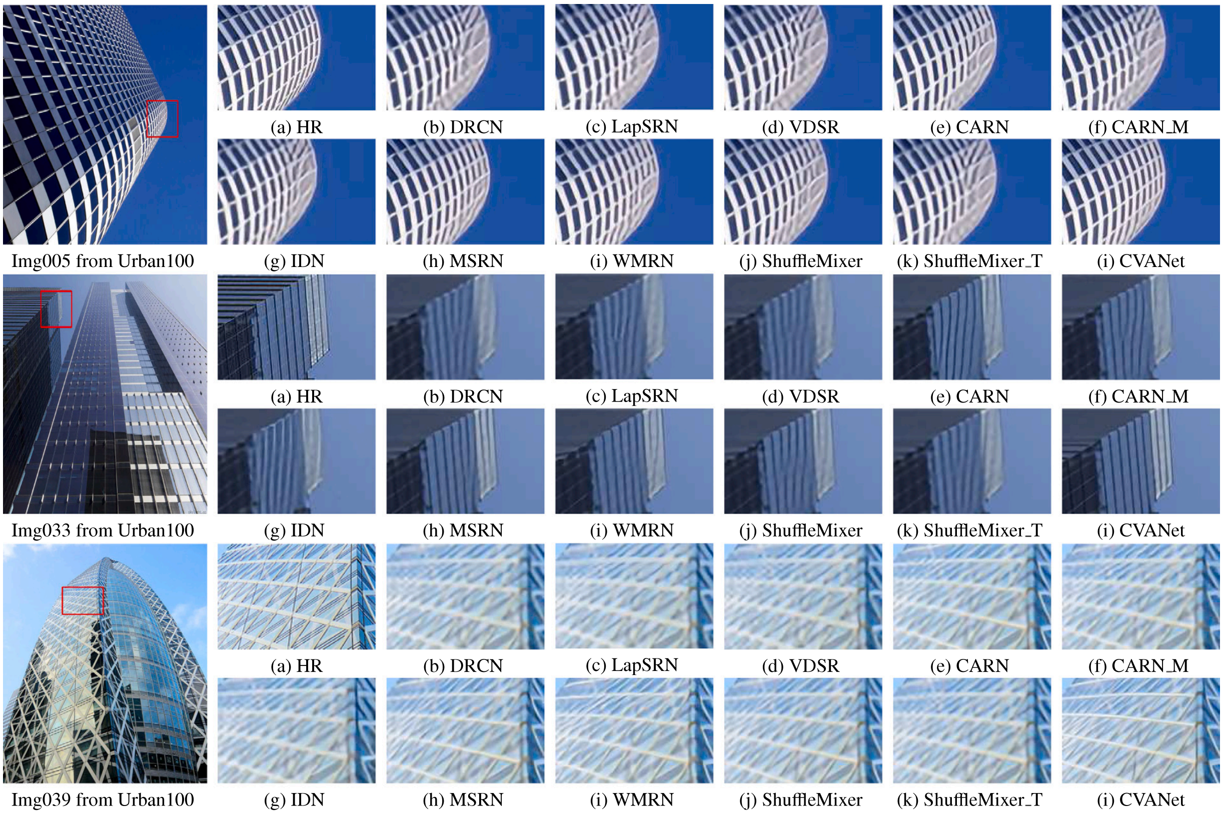


Fig. 8. Visual perception comparison results of different SR methods on the Urban100 dataset with ×4 SR images.

contains the attention information at the feature level. While our method mainly focuses on learning step-to-step attention information for a more powerful representation ability. Finally, we remove batch-normalization (BN) in our proposed FAM to adapt to super-resolution tasks and save computing resources. Inspired by SENet (Hu et al., 2020), we adapt the global average pooling instead of the original convolutional neural networks (CNNs). Moreover, we remove BN in CAM to decrease computational complexity. Inspired by SENet (Hu et al., 2020), we adopt the global average pooling instead of the original CNNs. Moreover, we remove BN in the CAM to decrease computational complexity. Non-local attention module (Chen et al., 2021) introduce a global self-attention mechanism, which allows each level to adaptively adjust its representation based on representative of the input information. But Non-local module consumes a considerable computational overhead, and it is widely used in high-level computer vision tasks. More significantly, the introduction of sparse self-attention reduces the calculation. Combined with FAM and CAM, our proposed PAM can provide multi-attention characteristics in SR. Furthermore, we add a skip connection between the input and output, which enables the model to learn shallow and deep information in PAM. We also remove BN to fit the SR task.

Difference to MSRN (Li et al., 2018). They include three main differences between MSRN (Li et al., 2018) and our CVANet. The first one is the basic convolutional module. MSRN (Li et al., 2018) introduces multi-scale at a coarse level, which extends the receptive field by extending the number of convolution kernels. However, this simple operation hinders the representational ability of CNNs. To address this issue, we introduce Res2Net (Gao et al., 2021) to image super-resolution for the first time. Res2Net (Gao et al., 2021) exploits the multi-scale potential at a more granular level and provide feature-level attention information with different receptive field to the following modules. The second one is there is no channel-wise attention module in MSRN (Li et al., 2018), so it is impossible for MSRN (Li et al., 2018) to exploit the channel's relationship. However, Li, Guo, and Loy (2022) demonstrated that the channel relationship is crucial for CNNs to exploit representative features. Therefore, we introduce channel attention with global average pooling to further use the relationship between channels. The third one is MSRN (Li et al., 2018) treats the feature maps extracted by previous layers equally, ignoring the pixel level relationship information. We use our proposed PA module to solve this issue. Experimental results on well-known datasets demonstrate that CVANet surpasses other methods.

Difference to RCAN (Zhang, Li, et al., 2018). There are two main differences between RCAN (Zhang, Li, et al., 2018) and our proposed CVANet. Although RCAN (Zhang, Li, et al., 2018) introduces a channel attention mechanism, it does not fully consider that the elements that make up the final SR image are pixels generated via features. Recent works (Ji et al., 2022; Zhao et al., 2022) has demonstrated that global information has great advantages in improving performance for low-level tasks. Our CVANet designs an efficient self-attention mechanism to fully extract the global features of low-resolution images and reduce the amount of computation. To refine the pixels forming SR, we introduce the PAM, and experiments prove that the introduced module can significantly improve the statistical results and visual effects.

Difference to SOTA methods. We summarize the main differences between existing SOTA methods and our CVANet. Existing SOTA methods usually simplify stacking a single module or only exploit one level of attention information, making these methods overly focus on local features, hindering the feature representational power of DCNNs. Our proposed method solves this issue by combining FAM, CAM, and PAM, which simulate human visual characteristics to refine the reconstruction steps. The simple improvement brings remarkable results and

Table 2
Statistic results of different modules for the implementation of ablation studies on the test datasets with ×2 SR images. These metrics represent the value of PSNR. (Optimal: red; Suboptimal: blue).

Method	Set 5	Set 14	B100	Urban 100
Baseline	37.86	33.51	32.13	31.88
-w/o F	37.97	33.70	32.20	32.26
-w/o C	38.01	33.54	32.21	32.18
-w/o P	37.97	33.68	32.20	32.22
-w/o FC	38.00	33.69	32.21	32.26
-w/o FP	38.00	33.62	32.19	32.18
-w/o CP	37.89	33.57	32.14	31.85
CVANet (full model)	38.05	33.73	32.21	32.26

forces the model to focus on different information in SR. Moreover, our proposed method is a generic framework that somebody can easily embed our proposed modules into their architecture depending on their needs.

4.4. Ablation analysis

We conducted the following ablation study using the widely used super-resolution dataset to evaluate the beneficial effects of each module in our CVANet on discriminatory performance, including (1) our CVANet without feature attention module (-w/o F), (2) our CVANet without channel attention module (-w/o C), (3) our CVANet without pixel attention module (-w/o P), (4) our CVANet without feature and channel attention modules (-w/o FC), (5) our CVANet without feature and pixel attention modules (-w/o FP), (6) our CVANet without channel and pixel attention modules (-w/o CP), (7) our CVANet with the full model.

From Table 2, the following observations can be summarized: (1) The three modules proposed can improve the performance of our CVANet, indicating that the proposed modules are effective in improving the performance of SR tasks. (2) -w/o CP can only slightly improve the performance of our CVANe as feature attention only provides a larger receptive field. (3) -w/o FC can significantly improve the performance of our CVANe as pixel attention refines the pixels utilized to form SR images. It is worth mentioning that in the ablation study, we conducted a total of 200 epochs. It is noteworthy that CVANet achieved the highest PSNR compared to the Baseline after 200 epochs. From Table 2, we designed each module to impact our CVANet positively. To sum up, our full model has the best PSNR scores in the four benchmarks (Bevilacqua et al., 2012; Huang et al., 2015; Martin et al., 2001; Zeyde et al., 2012).

4.5. Scalability

To illustrate that the concepts and modules we proposed are universal, we also conducted experiments based on the MSRN model and compared related methods, including SelfExSR (Huang et al., 2015), ESPCN (Shi et al., 2016), FSRCNN (Dong et al., 2016), VDSR (Kim et al., 2016a), DRCN (Kim et al., 2016b), LapSRN (Lai et al., 2017), and MSRN (Li et al., 2018). Notably, we performed only 800 epochs in the scalability study. The experimental results statistics in Table 3 show that the CVANet we designed can still significantly improve the model performance in the shallower model.

4.6. Limited applications

Image super-resolution reconstruction refers to using image processing methods to take a low-resolution image or sequence of images and process it to recover a high-resolution image. High resolution means that the image has a high pixel density, which provides more detail that is often critical in engineering practice and practical applications. In recent years, the technology has been gradually applied to satellite surveillance, underwater scenes, medical imaging, and other fields. Although our CVANet has better reconstruction performance for several standard datasets, our method has some limitations. The super-resolution results of our CVANet for × 2, × 3, and × 4 of low-light images, underwater images, and fog remote sensing images are shown in Fig. 9. Since our CVANet is only capable of reconstruction, it does not have the ability to correct for problems such as low brightness, color distortion and fog faced by images. Therefore, the reconstruction results of our CVANet for low-light, underwater, and fog remote sensing images are not satisfactory compared to the bilinear interpolation method.

5. Conclusion

To solve the issue that existing DCNN-based SISR methods cannot fully exploit the relationship between feature maps, channels, and pixels, we present a cascaded visual attention network for SISR to simulate human visual characteristics to refine the reconstruction steps, which can focus on details progressive like human beings and extracts the most representative features. Moreover, the proposed network is a multi-level attention network, which uses multiple attention information rather than single attention characteristics. Meanwhile, our CVANet is an end-to-end model, and the three attention modules can be easily embedded in other low-level computer vision tasks to improve performance. Extensively qualitative and quantitative experiments on several benchmarks illustrate that our proposed CVANet can significantly improve the validity and robustness of the baseline model.

Despite the obvious advantages of our CVANet, it also faces some limitations. On the one hand, too many attention models are introduced into the benchmark model, making it computationally more expensive. On the other hand, the complex network's robustness also faces some challenges. In the future, we will focus on designing lightweight network models, using Efficient Modules, and employing adaptive collaborative learning for the single image super-resolution.

Table 3
Performance of embedding our designed modules in the shallower model with ×2 SR images.

Method	Set5	Set14	BSDS100	Urban100
	PSRN/SSIM	PSRN/SSIM	PSRN/SSIM	PSRN/SSIM
BICUBIC (Dengwen, 2010)	33.69/0.9284	30.34/0.8675	29.57/0.8434	26.88/0.8438
SelfExSR (Huang et al., 2015)	36.60/0.9537	32.46/0.9051	31.20/0.8863	29.55/0.8983
ESPCN (Shi et al., 2016)	37.00/0.9559	32.75/0.9098	31.51/0.8939	29.87/0.9065
FSRCNN (Dong et al., 2016)	37.06/0.9554	32.76/0.9078	31.53/0.8912	29.88/0.9024
VDSR (Kim et al., 2016a)	37.53/0.9583	33.05/0.9107	31.92/0.8965	30.79/0.9157
DRCN (Kim et al., 2016b)	37.63/0.9584	33.06/0.9108	31.85/0.8947	30.76/0.9147
LapSRN (Lai et al., 2017)	37.52/0.9581	33.08/0.9109	31.80/0.8949	30.41/0.9112
MSRN (Li et al., 2018)	38.08/0.9605	33.74/0.9170	32.23/0.9013	32.22/0.9326
CVANet (ours)	38.19/0.9613	33.97/0.9210	32.33/0.9015	32.90/0.9353

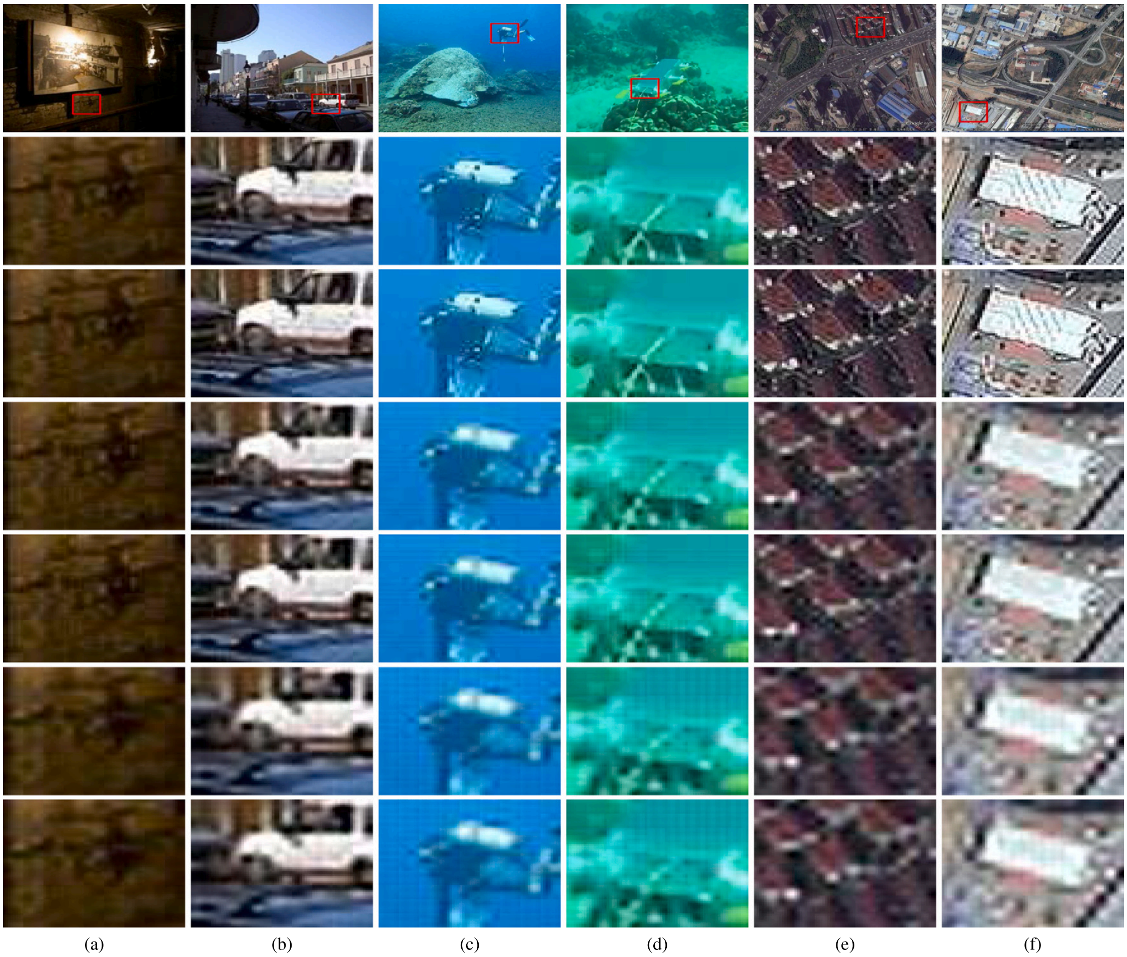


Fig. 9. Super-resolution results of our CVANet for several samples. From top to bottom are the original image, the × 2 SR image of bilinear interpolation, the × 2 SR image of CVANet, the × 3 SR image of bilinear interpolation, the × 3 SR image of CVANet, the × 4 SR image of bilinear interpolation, and the × 4 SR image of CVANet, respectively. Low-light images in (a) and (b) are sampled from the Zhang, Jin, Zhuang, Liang, and Li (2023) and Zhang, Wang, and Li (2022). Underwater images in (c) and (d) are sampled from the UIEB dataset (Li, Guo, et al., 2020). Fog remote sensing images in (e) and (f) are sampled from the Internet.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

This work was supported in part by the Natural Science Foundation of Henan Province, China under Grant 232300420428, and the National Natural Science Foundation of China under Grant 62171252.

References

Agustsson, Eirikur, & Timofte, Radu (2017). NTIRE 2017 challenge on single image super-resolution: Dataset and study. In *2017 IEEE conference on computer vision and pattern recognition workshops* (pp. 1122–1131).

Ahn, Namhyuk, Kang, Byungkon, & Sohn, Kyung-Ah (2018). Fast, accurate, and lightweight super-resolution with cascading residual network. In *Proceedings of the European conference on computer vision* (pp. 252–268).

Bevilacqua, Marco, Roumy, Aline, Guillemot, Christine, & Morel, Marie-Line Alberi (2012). Low-complexity single-image super-resolution based on nonnegative neighbor embedding. In *BMVC 2012 - Electronic proceedings of the British machine vision conference 2012*.

Caballero, Jose, Ledig, Christian, Aitken, Andrew, Acosta, Alejandro, Totz, Johannes, Wang, Zehan, et al. (2017). Real-time video super-resolution with spatio-temporal networks and motion compensation. In *2017 IEEE conference on computer vision and pattern recognition* (pp. 2848–2857).

Cao, Yang, Chandrasekar, A., Radhika, T., & Vijayakumar, V. (2023). Input-to-state stability of stochastic Markovian jump genetic regulatory networks. *Mathematics and Computers in Simulation*.

Chandrasekar, A., Radhika, T., & Zhu, Quanxin (2022a). Further results on input-to-state stability of stochastic Cohen–Grossberg BAM neural networks with probabilistic time-varying delays. *Neural Processing Letters*, 1–23.

Chandrasekar, A., Radhika, T., & Zhu, Quanxin (2022b). State estimation for genetic regulatory networks with two delay components by using second-order reciprocally convex approach. *Neural Processing Letters*, 1–19.

Chen, Tong, Liu, Haojie, Ma, Zhan, Shen, Qiu, Cao, Xun, & Wang, Yao (2021). End-to-end learnt image compression via non-local attention optimization and improved context modeling. *IEEE Transactions on Image Processing*, *30*, 3179–3191.

Chen, De-Lei, Zhang, Lei, & Huang, Hua (2022). Robust extraction and super-resolution of low-resolution flying airplane from satellite video. *IEEE Transactions on Geoscience and Remote Sensing*, *60*, 1–16.

Dai, Tao, Cai, Jianrui, Zhang, Yongbing, Xia, Shu-Tao, & Zhang, Lei (2019). Second-order attention network for single image super-resolution. In *2019 IEEE/CVF conference on computer vision and pattern recognition* (pp. 11057–11066).

Dengwen, Zhou (2010). An edge-directed bicubic interpolation algorithm. In *2010 3rd International Congress on Image and Signal Processing, Vol. 3* (pp. 1186–1189).

Dong, Chao, Loy, Chen Change, & Tang, Xiaoou (2016). Accelerating the super-resolution convolutional neural network. In *Computer vision–ECCV 2016: 14th European conference, Amsterdam, the Netherlands, October 11-14, 2016, Proceedings, Part II 14* (pp. 391–407).

Dumoulin, Vincent, Shlens, Jonathon, & Kudlur, Manjunath (2017). A learned representation for artistic style. In *5th International conference on learning representations, ICLR 2017 - conference track proceedings*.

Esmaeilzehi, Alireza, Ahmad, M. Omair, & Swamy, M. N. S. (2022). Ultralight-weight three-prior convolutional neural network for single image super resolution. *IEEE Transactions on Artificial Intelligence*, 1–15.

Fang, Faming, Li, Juncheng, & Zeng, Tieyong (2020). Soft-edge assisted network for single image super-resolution. *IEEE Transactions on Image Processing*, *29*, 4656–4668.

Fang, Jinsheng, Lin, Hanjiang, Chen, Xinyu, & Zeng, Kun (2022). A hybrid network of CNN and transformer for lightweight image super-resolution. In *2022 IEEE/CVF conference on computer vision and pattern recognition workshops (CVPRW)* (pp. 1102–1111).

Gao, Shang-Hua, Cheng, Ming-Ming, Zhao, Kai, Zhang, Xin-Yu, Yang, Ming-Hsuan, & Torr, Philip (2021). Res2Net: A new multi-scale backbone architecture. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *43*(2), 652–662.

Han, Jun, Zheng, Hao, Chen, Danny Z., & Wang, Chaoli (2022). STNet: An end-to-end generative framework for synthesizing spatiotemporal super-resolution volumes. *IEEE Transactions on Visualization and Computer Graphics*, *28*(1), 270–280.

He, Xiangyu, Mo, Zitao, Wang, Peisong, Liu, Yang, Yang, Mingyuan, & Cheng, Jian (2019). ODE-inspired network design for single image super-resolution. In *2019 IEEE/CVF conference on computer vision and pattern recognition* (pp. 1732–1741).

Hu, Jie, Shen, Li, Albanie, Samuel, Sun, Gang, & Wu, Enhua (2020). Squeeze-and-excitation networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *42*(8), 2011–2023.

Huang, Jia-Bin, Singh, Abhishek, & Ahuja, Narendra (2015). Single image super-resolution from transformed self-exemplars. In *2015 IEEE conference on computer vision and pattern recognition* (pp. 5197–5206).

Hui, Zheng, Wang, Xiumei, & Gao, Xinbo (2018). Fast and accurate single image super-resolution via information distillation network. In *2018 IEEE/CVF conference on computer vision and pattern recognition* (pp. 723–731).

Ji, Li, Zhu, Qinghui, Zhang, Yongqin, Yin, Juanjuan, Wei, Ruyi, Xiao, Jinsheng, et al. (2022). Cross-domain heterogeneous residual network for single image super-resolution. *Neural Networks*, *149*, 84–94.

Jiang, Junjun, Yu, Yi, Wang, Zheng, Tang, Suhua, Hu, Ruimin, & Ma, Jiayi (2020). Ensemble super-resolution with a reference dataset. *IEEE Transactions on Cybernetics*, *50*(11), 4694–4708.

Kim, Jiwon, Lee, Jung Kwon, & Lee, Kyoung Mu (2016a). Accurate image super-resolution using very deep convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1646–1654).

Kim, Jiwon, Lee, Jung Kwon, & Lee, Kyoung Mu (2016b). Deeply-recursive convolutional network for image super-resolution. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1637–1645).

Kim, Bumsoo, Mun, Jonghwan, On, Kyoung-Woon, Shin, Minchul, Lee, Junhyun, & Kim, Eun-Sol (2022). MSTR: Multi-scale transformer for end-to-end human-object interaction detection. In *2022 IEEE/CVF conference on computer vision and pattern recognition* (pp. 19556–19565).

Kingma, Diederik P., & Ba, Jimmy Lei (2015). Adam: A method for stochastic optimization. In *3rd International conference on learning representations, ICLR 2015 - conference track proceedings*.

Lai, Wei-Sheng, Huang, Jia-Bin, Ahuja, Narendra, & Yang, Ming-Hsuan (2017). Deep laplacian pyramid networks for fast and accurate super-resolution. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 624–632).

Lan, Rushi, Sun, Long, Liu, Zhenbing, Lu, Huimin, Su, Zhixun, Pang, Cheng, et al. (2021). Cascading and enhanced residual networks for accurate single-image super-resolution. *IEEE Transactions on Cybernetics*, *51*(1), 115–125.

Lei, Haijun, Tian, Zhihui, Xie, Hai, Zhao, Benjian, Zeng, Xianlu, Cao, Jiuwen, et al. (2023). LAC-GAN: Lesion attention conditional GAN for ultra-widefield image synthesis. *Neural Networks*, *158*, 89–98.

Lei, Jianjun, Zhang, Zhe, Fan, Xiaoting, Yang, Bolan, Li, Xinxin, Chen, Ying, et al. (2021). Deep stereoscopic image super-resolution via interaction module. *IEEE Transactions on Circuits and Systems for Video Technology*, *31*(8), 3051–3061.

Li, Chongyi, Cong, Runmin, Kwong, Sam, Hou, Junhui, Fu, Huazhu, Zhu, Guopu, et al. (2021). ASIF-net: Attention steered interweave fusion network for RGB-D salient object detection. *IEEE Transactions on Cybernetics*, *51*(1), 88–100.

Li, Juncheng, Fang, Faming, Mei, Kangfu, & Zhang, Guixu (2018). Multi-scale residual network for image super-resolution. In *Proceedings of the European conference on computer vision* (pp. 517–532).

Li, Chongyi, Guo, Chunle, & Loy, Chen Change (2022). Learning to enhance low-light image via zero-reference deep curve estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *44*(8), 4225–4238.

Li, Chongyi, Guo, Chunle, Ren, Wenqi, Cong, Runmin, Hou, Junhui, Kwong, Sam, et al. (2020). An underwater image enhancement benchmark dataset and beyond. *IEEE Transactions on Image Processing*, *29*, 4376–4389.

Li, Tao, Lin, Hongwei, Dong, Xiucheng, & Zhang, Xiaohua (2020). Depth image super-resolution using correlation-controlled color guidance and multi-scale symmetric network. *Pattern Recognition*, *107*, Article 107513.

Li, Xiang, Wang, Wenhai, Hu, Xiaolin, & Yang, Jian (2019). Selective kernel networks. In *2019 IEEE/CVF conference on computer vision and pattern recognition* (pp. 510–519).

Li, Xiang, Wang, Wenhai, Hu, Xiaolin, & Yang, Jian (2020). Selective kernel networks. In *2019 IEEE/CVF conference on computer vision and pattern recognition* (pp. 510–519).

Luo, Wenjie, Li, Yujia, Urtasun, Raquel, & Zemel, Richard (2016). Understanding the effective receptive field in deep convolutional neural networks. *Advances in Neural Information Processing Systems*, *29*.

Martin, D., Fowlkes, C., Tal, D., & Malik, J. (2001). A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In *Proceedings eighth IEEE international conference on computer vision, vol. 2* (pp. 416–423).

Qin, Jiayi, Chen, Lihui, Jeon, Seunggil, & Yang, Xiaomin (2023). Progressive interaction-learning network for lightweight single-image super-resolution in industrial applications. *IEEE Transactions on Industrial Informatics*, *19*(2), 2183–2191.

Radhika, T., Chandrasekar, A., Vijayakumar, V., & Zhu, Quanxin (2023). Analysis of Markovian jump stochastic Cohen–Grossberg BAM neural networks with time delays for exponential input-to-state stability. *Neural Processing Letters*, 1–18.

Rakkiyappan, Rajan, Chandrasekar, Arunachalam, & Cao, Jinde (2014). Passivity and passification of memristor-based recurrent neural networks with additive time-varying delays. *IEEE Transactions on Neural Networks and Learning Systems*, *26*(9), 2043–2057.

Ran, Ran, Deng, Liang-Jian, Jiang, Tai-Xiang, Hu, Jin-Fan, Chanussot, Jocelyn, & Vivone, Gemine (2023). GuidedNet: A general CNN fusion framework via high-resolution guidance for hyperspectral image super-resolution. *IEEE Transactions on Cybernetics*, 1–14.

Ren, Zeyu, Kong, Xiangyu, Zhang, Yudong, & Wang, Shuihua (2023). UKSSL: Underlying knowledge based semi-supervised learning for medical image classification. *IEEE Open Journal of Engineering in Medicine and Biology*, 1–8. http://dx.doi.org/10.1109/OJEMB.2023.3305190.

Ren, Zeyu, Wang, Shuihua, & Zhang, Yudong (2023a). Weakly supervised machine learning. *CAAI Transactions on Intelligence Technology*, 549–580.

Ren, Zeyu, Wang, Shuihua, & Zhang, Yudong (2023b). Weakly supervised machine learning. *CAAI Transactions on Intelligence Technology*.

Shi, Wenzhe, Caballero, Jose, Huszár, Ferenc, Totz, Johannes, Aitken, Andrew P., Bishop, Rob, et al. (2016). Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. In *2016 IEEE conference on computer vision and pattern recognition* (pp. 1874–1883).

Song, Zijiang, & Zhong, Baojiang (2022). A lightweight local-global attention network for single image super-resolution. In *Proceedings of the Asian conference on computer vision* (pp. 4395–4410).

Sun, Long, Liu, Zhenbing, Sun, Xiyan, Liu, Licheng, Lan, Rushi, & Luo, Xiaonan (2021). Lightweight image super-resolution via weighted multi-scale residual network. *IEEE/CAA Journal of Automatica Sinica*, *8*(7), 1271–1280.

Sun, Long, Pan, Jinshan, & Tang, Jinhui (2022). ShuffleMixer: An efficient ConvNet for image super-resolution. *Advances in Neural Information Processing Systems*.

Tai, Ying, Yang, Jian, & Liu, Xiaoming (2017). Image super-resolution via deep recursive residual network. In *2017 IEEE conference on computer vision and pattern recognition* (pp. 2790–2798).

Tamil Thendral, Marimuthu, Ganesh Babu, Thiruvannamalai Radhakrishnan, Chandrasekar, Arunachalam, & Cao, Yang (2022). Synchronization of Markovian jump neural networks for sampled data control systems with additive delay components: Analysis of image encryption technique. *Mathematical Methods in the Applied Sciences*.

Tian, Chunwei, Xu, Yong, Zuo, Wangmeng, Zhang, Bob, Fei, Lunke, & Lin, Chia-Wen (2021). Coarse-to-fine CNN for image super-resolution. *IEEE Transactions on Multimedia*, *23*, 1489–1502.

Tian, Chunwei, Yuan, Yixuan, Zhang, Shichao, Lin, Chia-Wen, Zuo, Wangmeng, & Zhang, David (2022). Image super-resolution with an enhanced group convolutional neural network. *Neural Networks*, *153*, 373–385.

Wang, Longguang, Dong, Xiaoyu, Wang, Yingqian, Ying, Xinyi, Lin, Zaiping, An, Wei, et al. (2021). Exploring sparsity in image super-resolution for efficient inference. In *2021 IEEE/CVF conference on computer vision and pattern recognition* (pp. 4915–4924).

Wang, Xinya, Ma, Jiayi, Jiang, Junjun, & Zhang, Xiao-Ping (2022). Dilated projection correction network based on autoencoder for hyperspectral image super-resolution. *Neural Networks*, *146*, 107–119.

Wang, Yan, Su, Tongtong, Li, Yusen, Cao, Jiuwen, Wang, Gang, & Liu, Xiaoguang (2022). Ddistill-SR: Reparameterized dynamic distillation network for lightweight image super-resolution. *IEEE Transactions on Multimedia*, 1–13.

Wu, Huapeng, Zou, Zhengxia, Gui, Jie, Zeng, Wen-Jun, Ye, Jieping, Zhang, Jun, et al. (2021). Multi-grained attention networks for single image super-resolution. *IEEE Transactions on Circuits and Systems for Video Technology*, *31*(2), 512–522.

Xin, Jingwei, Li, Jie, Jiang, Xinrui, Wang, Nannan, Huang, Heng, & Gao, Xinbo (2022). Wavelet-based dual recursive network for image super-resolution. *IEEE Transactions on Neural Networks and Learning Systems*, *33*(2), 707–720.

Yan, Yanyang, Ren, Wenqi, Hu, Xiaobin, Li, Kun, Shen, Haifeng, & Cao, Xiaochun (2021). SRGAT: Single image super-resolution with graph attention network. *IEEE Transactions on Image Processing*, *30*, 4905–4918.

Zeyde, Roman, Elad, Michael, & Protter, Matan (2012). On single image scale-up using sparse-representations. *Lecture Notes in Computer Science*, *6920 LNCS*, 711–730.

Zhang, Yudong, Deng, Lijia, Zhu, Hengde, Wang, Wei, Ren, Zeyu, Zhou, Qinghua, et al. (2023). Deep learning in food category recognition. *Information Fusion*, Article 101859.

Zhang, Weidong, Jin, Songlin, Zhuang, Peixian, Liang, Zheng, & Li, Chongyi (2023). Underwater image enhancement via piecewise color correction and dual prior optimized contrast enhancement. *IEEE Signal Processing Letters*, *30*, 229–233.

Zhang, Yulun, Li, Kunpeng, Li, Kai, Wang, Lichen, Zhong, Bineng, & Fu, Yun (2018). Image super-resolution using very deep residual channel attention networks. In *Proceedings of the European conference on computer vision* (pp. 286–301).

Zhang, Weidong, Li, Zexu, Sun, Hai-Han, Zhang, Qiang, Zhuang, Peixian, & Li, Chongyi (2022). SSTNet: Spatial, spectral, and texture aware attention network using hyperspectral image for corn variety identification. *IEEE Geoscience and Remote Sensing Letters*, *19*, 1–5.

Zhang, Yongbing, Liu, Siyuan, Dong, Chao, Zhang, Xinfeng, & Yuan, Yuan (2020). Multiple cycle-in-cycle generative adversarial networks for unsupervised image super-resolution. *IEEE Transactions on Image Processing*, *29*, 1101–1112.

Zhang, Yulun, Tian, Yapeng, Kong, Yu, Zhong, Bineng, & Fu, Yun (2018). Residual dense network for image super-resolution. In *2018 IEEE/CVF conference on computer vision and pattern recognition* (pp. 2472–2481).

Zhang, Weidong, Wang, Yudong, & Li, Chongyi (2022). Underwater image enhancement by attenuated color channel correction and detail preserved contrast enhancement. *IEEE Journal of Oceanic Engineering*, *47*(3), 718–735.

Zhang, Weidong, Zhuang, Peixian, Sun, Hai-Han, Li, Guohou, Kwong, Sam, & Li, Chongyi (2022). Underwater image enhancement via minimal color loss and locally adaptive contrast enhancement. *IEEE Transactions on Image Processing*, *31*, 3997–4010.

Zhao, Wenyi, Yang, Huihua, Pan, Xipeng, & Li, Lingqiao (2021). S^2-aware network for visual recognition. *Signal Processing: Image Communication*, *99*, Article 116458.

Zhao, Wenyi, Yang, Lu, Zhang, Weidong, Tian, Yongqin, Jia, Wenhe, Li, Wei, et al. (2023). Learning what and where to learn: A new perspective on self-supervised learning. *IEEE Transactions on Circuits and Systems for Video Technology*, 1. http://dx.doi.org/10.1109/TCSVT.2023.3298937.

Zhao, Wenyi, Zhang, Weidong, Pan, Xipeng, Zhuang, Peixian, Xie, Xiwang, Li, Lingqiao, et al. (2022). LESSL: Can LEGO sampling and collaborative optimization contribute to self-supervised learning? *Information Sciences*, *615*, 475–490.

Zhuang, Peixian, Wu, Jiamin, Porikli, Fatih, & Li, Chongyi (2022). Underwater image enhancement with hyper-Laplacian reflectance priors. *IEEE Transactions on Image Processing*, *31*, 5442–5455.