**ARTICLE**

# Pyramid Separable Channel Attention Network for Single Image Super-Resolution

**Congcong Ma[1,3], Jiaqi Mi[2], Wanlin Gao[1,3] and Sha Tao[1,3,*]**

[1]College of Information and Electrical Engineering, China Agricultural University, Beijing, 100083, China

[2]College of Artificial Intelligence, Nankai University, Tianjin, 300350, China

[3]Key Laboratory of Agricultural Informatization Standardization, Ministry of Agriculture and Rural Affairs, China Agricultural University, Beijing, 100083, China

*Corresponding Author: Sha Tao. Email: taosha20070608@163.com

## ABSTRACT

Single Image Super-Resolution (SISR) technology aims to reconstruct a clear, high-resolution image with more information from an input low-resolution image that is blurry and contains less information. This technology has significant research value and is widely used in fields such as medical imaging, satellite image processing, and security surveillance. Despite significant progress in existing research, challenges remain in reconstructing clear and complex texture details, with issues such as edge blurring and artifacts still present. The visual perception effect still needs further enhancement. Therefore, this study proposes a Pyramid Separable Channel Attention Network (PSCAN) for the SISR task. This method designs a convolutional backbone network composed of Pyramid Separable Channel Attention blocks to effectively extract and fuse multi-scale features. This expands the model's receptive field, reduces resolution loss, and enhances the model's ability to reconstruct texture details. Additionally, an innovative artifact loss function is designed to better distinguish between artifacts and real edge details, reducing artifacts in the reconstructed images. We conducted comprehensive ablation and comparative experiments on the Arabidopsis root image dataset and several public datasets. The experimental results show that the proposed PSCAN method achieves the best-known performance in both subjective visual effects and objective evaluation metrics, with improvements of 0.84 in Peak Signal-to-Noise Ratio (PSNR) and 0.017 in Structural Similarity Index (SSIM). This demonstrates that the method can effectively preserve high-frequency texture details, reduce artifacts, and have good generalization performance.

## KEYWORDS

Deep learning; single image super-resolution; artifacts; texture details

## 1 Introduction

Single Image Super-Resolution (SISR) aims to reconstruct high-resolution images from low-resolution inputs [1,2]. This technology is widely used in fields such as medical imaging [3], satellite image processing [4], surveillance systems, and entertainment. Despite the significant progress made by deep learning-based methods in recent years, reconstructed images often exhibit various artifacts,

such as blurriness, ringing effects, and ghosting. These artifacts severely affect the visual quality and practical value of the images.

Existing SISR research generally relies on fundamental digital image processing techniques for reconstruction. Common methods mainly include interpolation-based methods [5], degradation model-based methods [6,7], and learning-based methods [8]. Interpolation-based methods are straightforward to implement [9], but they often result in output images with ringing artifacts and blurred edges. Degradation model-based methods extract key information from low-resolution (LR) images [10] and constrains the generation of super-resolution (SR) images using prior knowledge of the unknown high-resolution (HR) images. However, the degradation process in real-world images is complex and noisy, making practical application challenging. Learning-based methods [11] use large amounts of training data to learn the mapping relationship between LR and HR images, thereby predicting the corresponding SR images. These methods do not require extensive prior knowledge and, with recent advances in hardware performance, deep learning-based super-resolution techniques [12–15] have achieved better reconstruction quality. Although many studies, such as Very Deep network for Super-Resolution (VDSR) [16] and Super-Resolution Generative Adversarial Network (SRGAN) [17], have improved image quality by introducing deeper network structures, adversarial loss, and perceptual loss, these methods still have limitations. For instance, deep networks can suffer from gradient vanishing issues during training, and while adversarial loss can help reduce certain artifacts, it may also introduce new ones.

This paper proposes a Pyramid Separable Channel Attention Network (PSCAN) based on convolutional neural networks for single image super-resolution. The main innovations and contributions of this study are as follows:

- **Design of the Pyramid Separable Convolutional Channel Attention Block (PSCAB):** This module effectively extracts image features, where convolutional kernels of different sizes in the pyramid separable convolution allow the model to obtain multi-scale receptive fields, thereby enhancing the model's ability to restore high-frequency details.
- **Optimization of the Loss Function:** An artifact loss factor is introduced based on the L1 loss function to distinguish high-frequency artifacts from details, reduce artifacts in the reconstructed images, and improve their overall visual quality.
- **Extensive Experimental Validation:** Comprehensive experiments were conducted on the model plant Arabidopsis root image dataset and multiple standard public datasets for SISR tasks. Notably, on the challenging "Root" dataset, the super-resolution images generated by PSCAN achieved state-of-the-art results in both visual quality and objective evaluation metrics (PSNR and SSIM). The results demonstrate that the proposed PSCAN model can effectively reduce artifacts in super-resolution images, enhance the visual quality and detail retention of reconstructed images, and exhibit good generalization ability.

The remainder of this paper is organized as follows: Section 2 reviews related work, Section 3 provides a detailed description of the proposed method, Section 4 introduces the experimental setup, Section 5 presents the experimental results, and Section 6 discusses the findings and concludes the paper.

## 2 Related Works

In the field of Single Image Super-Resolution, numerous studies aim to improve image quality and reduce artifacts. This section reviews deep learning-based SISR methods in four parts.

### 2.1 CNN-Based Methods

The advent of deep learning has revolutionized the SISR domain. Dong et al. proposed the Super-Resolution Convolutional Neural Network (SRCNN) [18], which significantly improved performance by learning an end-to-end mapping from low-resolution to high-resolution images. The VDSR model enhanced the learning capability by deepening the network, but it performed poorly in retaining high-frequency details, often producing blurred images. To overcome the limitations of early deep learning models, researchers have explored multi-scale feature extraction techniques. For instance, the Residual Dense Network (RDN) [19] and the Dual Regression Network (DRN) [20] integrate hierarchical features at different scales to enhance texture and detail reconstruction. These models demonstrate that leveraging multi-scale information can effectively improve the clarity and sharpness of super-resolved images.

### 2.2 Attention Mechanism-Based Methods

Attention mechanisms have achieved significant success in recent deep learning research. The Residual Channel Attention Network (RCAN) [21] leverages channel attention mechanisms to enhance feature representation through Residual Channel Attention Blocks. The Hierarchical Attention Network (HAN) [22] introduces layer attention modules and channel-spatial attention modules to jointly learn feature information across layers, channels, and positions, improving the model's feature extraction capabilities. The Second-Order Attention Network (SAN) [23] incorporates second-order attention mechanisms, further enhancing feature extraction, particularly in handling complex textures and details. Incorporating attention mechanisms into SISR tasks effectively captures both global and local information, improving reconstruction performance.

### 2.3 GAN-Based Methods

Generative Adversarial Networks (GANs) [24], introduced by Goodfellow et al., have brought significant advancements in image generation and super-resolution. SRGAN [17], a pioneering work applying adversarial learning to SISR, facilitated the generation of perceptually sharper and more visually appealing images. SRGAN and its enhanced version, Enhanced Super-Resolution Generative Adversarial Networks (ESRGAN) [25], demonstrate that combining adversarial loss and perceptual loss can significantly reduce artifacts like ringing and enhance high-frequency details. Unfolding Super-Resolution Generative Adversarial Network (USRGAN) [26] improves model performance by embedding degradation constraints and prior knowledge, while SPSR [27] uses image gradient maps to guide super-resolution reconstruction, alleviating structural distortions in SR images. Content-aware Local GAN (CALGAN) [28] aims to achieve photorealistic super-resolution at the level of photographs. However, due to the aggressive nature of adversarial training, these methods sometimes introduce unnatural textures and other types of artifacts. Additionally, adversarial training can be unstable, requiring fine-tuning of hyperparameters and longer training times.

### 2.4 Transformer-Based Methods

Transformers have achieved significant success in natural language processing and have been introduced to SISR tasks to improve the quality of reconstructed images. Image Processing Transformer (IPT) [29] is a large-scale pre-trained Transformer model suitable for multiple image processing tasks, including SISR. Pre-trained on large-scale data, IPT demonstrates strong transfer learning capabilities and superior performance. Swin transformer for Image Restoration (SwinIR) [30], based on the Swin Transformer structure, performs well in various image restoration tasks, especially in SISR, achieving

significant performance improvements through efficient sliding window mechanisms and hierarchical feature extraction. However, Transformers typically require large amounts of data and computational resources for pre-training and fine-tuning, posing challenges in terms of computational cost and data requirements in practical applications.

Building on these advancements, our approach aims to enhance feature extraction capabilities, improve the visual quality of super-resolved images, and effectively reduce unwanted artifacts. The following sections will provide a detailed introduction to the PSCAN method and validate its effectiveness through extensive experiments.

## 3 Method

### 3.1 Overall Structure of PSCAN

The input to the PSCAN network is the LR image, which undergoes feature extraction and upsampling to generate a high-resolution SR image. The overall structure of the PSCAN model is illustrated in Fig. 1. The PSCAN primarily consists of convolutional layers, stacked residual groups, skip connections, and upsampling modules. Each residual group comprises stacked Pyramid Separable Channel Attention Blocks (**PSCAB**s), convolutional layers, and skip connections.
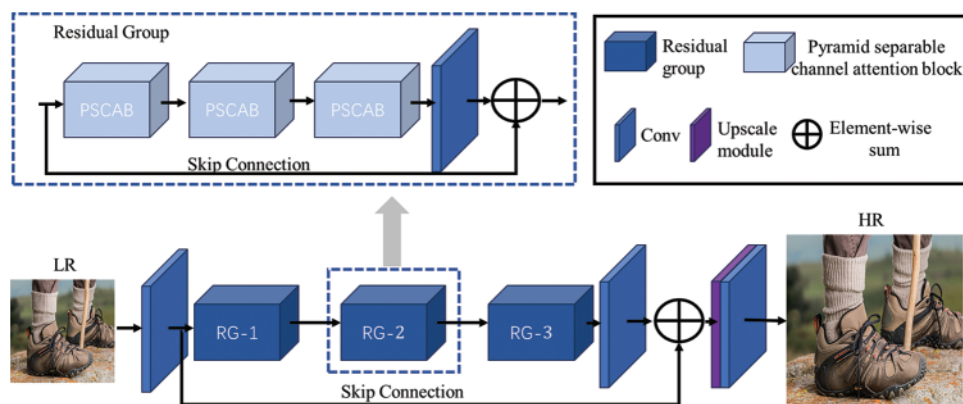


**Figure 1:** Overview of the PSCAN

### 3.2 Pyramid Separable Channel Attention Block

The detailed structure of PSCAB is illustrated in Fig. 2. The input feature maps from the upper layer sequentially pass through the pyramid separable convolution units (PSCU) and channel attention modules [21]. The output feature maps then undergo element-wise addition with the input from the upper layer through skip connections. The resulting feature maps serve as the output of PSCAB.

The channel attention mechanism assigns different weights to each channel of the feature map. However, conventional convolution operations do not treat channels as independent entities for computation; instead, they compute convolution simultaneously across all channels with the convolutional kernel. This misalignment with the fundamental principle of treating each channel independently in the channel attention mechanism may hinder its optimal effectiveness. To address this inconsistency, we attempted to introduce separable convolution, which independently processes individual channels. In addition, to enable the model to obtain different scales of receptive fields, we designed convolutional kernels of sizes $3 \times 3$, $5 \times 5$, and $7 \times 7$, respectively, ultimately forming pyramid

separable convolution units, as illustrated in Fig. 2. The pyramid separable channel attention block enlarges the model's receptive field, effectively extracting and integrating multi-scale image features, thereby enhancing the reconstruction capability of high-frequency details.
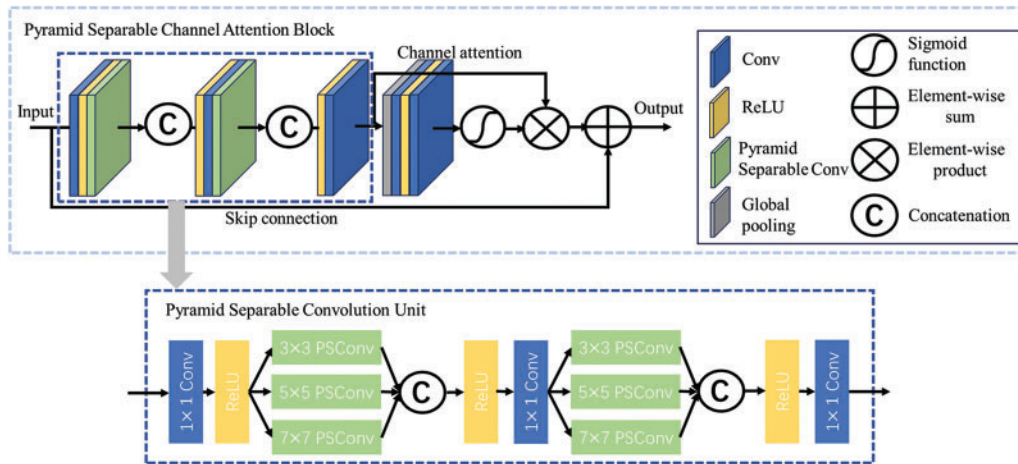


**Figure 2:** The structure of the PSCAB and PSCU

### 3.3 Artifact Loss Function

In image super-resolution tasks, artifacts commonly appear as edge ringing, halo effects, blocking artifacts, texture distortion, and noise artifacts. These artifacts significantly degrade the visual quality of the images. Additionally, artifacts can distort high-frequency details in super-resolved images, leading to the loss of critical information. They severely impact the overall quality of the images, undermining their natural appearance and the authenticity of fine details. Therefore, reducing or eliminating artifacts is crucial for enhancing image quality, particularly in super-resolution tasks. To minimize the generation of artifacts in reconstructed images, this paper proposes an artifact loss factor constructed based on the local residuals between the SR image and the HR image, aiming to discern artifacts from genuine details and enhance the reconstruction quality of the SR images. The artifact loss factor, along with the commonly used L1 loss in SISR tasks, constitutes the artifact loss function, as depicted in Eq. (1). The L1 loss is employed to measure the pixel-wise similarity between the SR image and the HR image, while the artifact loss factor aims to differentiate high-frequency details from artifacts.

$$Loss = \left\| H_{PSCAN} \left( I^{LR} \right) - I^{HR} \right\|_1 + \mu loss_{\text{artifact}} \tag{1}$$

where $I^{HR}$ represents the high-resolution image, and $H_{PSCAN} \left( I^{LR} \right)$ denotes the generated super-resolution image by the PSCAN, $\mu$ is a hyperparameter.

$loss_{\text{artifact}}$ is defined as the Kullback-Leibler (KL) divergence between the distributions of the residual image, obtained by subtracting the SR image from the HR image, and the edge information image extracted by applying a Gaussian high-pass filter to the HR image, as detailed in Eq. (2). The former, the residual image, contains mixed high-frequency artifacts and edge information, while the latter contains edge information from the HR image. We believe that the local dispersion of pixel areas in edge details differs from that in artifact regions. Therefore, to remove artifacts as much as possible while retaining edge details, it's essential to minimize the distance between the residual image and the HR edge image, making the distributions of these two images as similar as possible. Ultimately, the

Kullback-Leibler (KL) divergence is chosen to measure the closeness between these distributions. A smaller KL divergence indicates a higher similarity between the two distributions.

$$loss_{\text{artifact}} = KL\left(\left(I^{HR} - H_{PSCAN}\left(I^{LR}\right)\right) || GHPF\left(I^{HR}\right)\right) \tag{2}$$

The transfer function of the Gaussian high-pass filter is:

$$GHPF\left(I_{ij}^{HR}\right) = 1 - e^{-\frac{D^2(i,j)}{2D_0^2}} \tag{3}$$

$D_0$ is the distance from the frequency rectangular center to the cutoff frequency. It works better when set to 30. The Gaussian high-pass filter enhances details and edge information, effectively emphasizing small objects and fine lines.

Kullback-Leibler (KL) divergence is defined as:

$$KL(P||Q) = \sum_{ij} P_{ij} \log \frac{P_{ij}}{Q_{ij}} \tag{4}$$

## 4 Experiment Setting

### 4.1 Experimental Environment

Our experiments are conducted on the graphics processing units (GPUs) of NVIDA GeForce RTX 3060Ti with 8 GB graphics memory size, 14 GHz memory clock, bit width is 256 bits. In addition, the processor model of the computer is i7-12700K, the memory size is 32 GB, and the operating system is Window 10. The model implementation is based on TensorFlow 2.0 framework, Integrated Development environment (IDE) is PyCharm. The main toolkits used are numpy, random, glob, imageio, math, time, os, etc. The main programming language used is Python 3.7.

### 4.2 Hyperparameters

In the model training process, a batch size of 4 was used with the Adam optimizer, initialized with a learning rate of 0.0002. The hyperparameter μ in the loss function was set to 0.1. To ensure the reliability of model evaluation, 5-fold cross-validation technique was employed in the experiments.

### 4.3 Evaluation Metrics

Image super-resolution performance can be evaluated from both subjective visual effects and objective metrics. Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index (SSIM) are two fundamental metrics used to measure the quality of reconstructed images in super-resolution tasks.

PSNR is the ratio of the maximum power of a signal to the power of noise in an image, used to measure the quality of a reconstructed image that has been compressed, typically expressed in decibels (dB). Generally, a higher PSNR value indicates better image quality. The equation is as follows:

$$PSNR = 10 \times log_{10}\left(\frac{max\ I^2}{MSE}\right) = 20 \times log_{10}\left(\frac{max\ I}{\sqrt{MSE}}\right) \tag{5}$$

Here, *max I* is the maximum possible pixel value of the image. *MSE* is the Mean Squared Error between the original image and the reconstructed image.

The SSIM ranges from 0 to 1. A higher SSIM value indicates less image distortion and better image quality, as represented by Eq. (6).

$$SSIM\,(x,y) = \frac{\left(2\mu_x\mu_y + C_1\right)\left(2\sigma_{xy} + C_2\right)}{\left(\mu_x^2 + \mu_y^2 + C_1\right)\left(\sigma_x^2 + \sigma_y^2 + C_2\right)} \tag{6}$$

Here, $x$ and $y$ are the two images being compared. $\mu_x$ and $\mu_y$ represent the mean values of pixels in images $x$ and $y$, respectively. $\sigma_x^2$ and $\sigma_y^2$ denote the variances of pixels in images $x$ and $y$, respectively. $\sigma_{xy}$ denotes the covariance of pixels between images $x$ and $y$. $C_1$ and $C_2$ are constants to stabilize the computation and prevent division by zero. $C_1 = (k_1L)^2$ and $C_2 = (k_2L)^2$ where $L$ represents the maximum pixel value range (e.g., 255), and $k_1$ and $k_2$ are constants.

### 4.4 Datasets

Similar to other seminal studies, we conducted extensive comparative experiments on the standard publicly available DIV2K dataset to validate the overall performance of the proposed PSCAN model. We use 800 training images from DIV2K dataset [21] as training set. We conducted testing using 100 test images from the DIV2K dataset and four standard benchmark datasets, namely Set5 [31], Set14 [32], Urban100 [33], and Manga109 [34].

The DIV2K (Diverse 2K) dataset is a high-quality image super-resolution dataset widely utilized for training and evaluating image super-resolution algorithms. It encompasses various scenes, including natural landscapes, urban environments, and indoor settings. The images are rich in content, diverse in detail, and finely detailed, providing an excellent foundation for super-resolution tasks in both training and testing. These images have a resolution of 2K, and corresponding low-resolution images are obtained using Bicubic downsampling by a factor of four.

Additionally, we utilized a self-constructed "Root" dataset to more intuitively demonstrate the model's unique advantages in high-frequency detail restoration and artifact removal. Root images contain numerous fine roots and branching structures, characterized by rich details and complex textures. Due to these characteristics, they are prone to generating artifacts during the super-resolution process [35–37]. Therefore, they serve as an ideal dataset for evaluating the SISR model's capability to recover high-frequency details.

"Root": This dataset comprises 2500 images of Arabidopsis thaliana root systems, each sized at $256 \times 256$ pixels. Low-resolution images were derived using Bicubic downsampling by a factor of four, and the dataset was split into training and test sets at a ratio of 4:1.

## 5 Results

### 5.1 Overall Experimental Results

Similar to other seminal studies, we conducted extensive experiments on publicly available SISR task datasets to intuitively demonstrate the overall performance of the proposed method. In Fig. 3 and Table 1, we present the subjective visual comparison of the super-resolved images and the quantitative results in terms of PSNR and SSIM metrics, respectively.
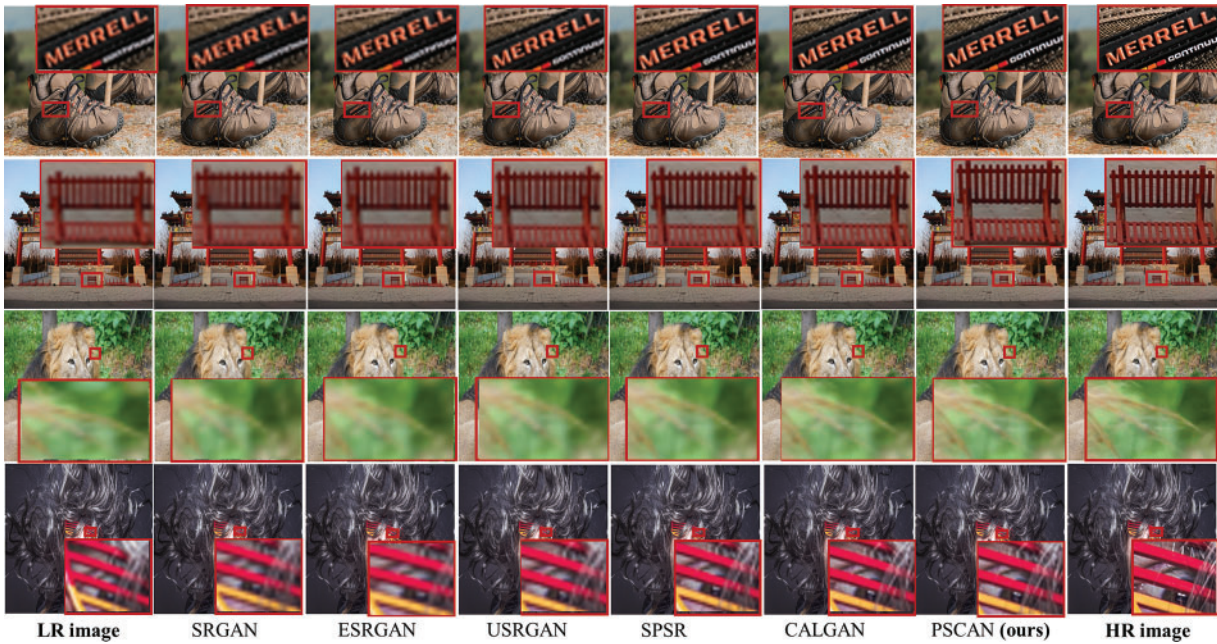
**Figure 3:** Visualization results of various SISR methods on the DIV2K test set. There are four groups of samples arranged from top to bottom, labeled "Shoes," "Building," "Lion," and "Person." Red boxes highlight and magnify local details in the super-resolution images

**Table 1:** PSNR and SSIM of various SISR models

| Methods | DIV2K | | Set5 | | Set14 | | Urban100 | | Manga109 | |
|---|---|---|---|---|---|---|---|---|---|---|
| | PSNR | SSIM | PSNR | SSIM | PSNR | SSIM | PSNR | SSIM | PSNR | SSIM |
| ANTIALIAS | 24.296 | 0.619 | 26.318 | 0.623 | 23.413 | 0.541 | 20.669 | 0.462 | 19.643 | 0.479 |
| NEAREST | 22.621 | 0.532 | 23.934 | 0.544 | 22.037 | 0.484 | 19.669 | 0.414 | 18.445 | 0.441 |
| BILINEAR | 23.436 | 0.561 | 25.174 | 0.581 | 22.684 | 0.492 | 20.074 | 0.409 | 18.954 | 0.437 |
| BICUBIC | 24.043 | 0.604 | 25.969 | 0.615 | 23.212 | 0.534 | 20.497 | 0.450 | 19.438 | 0.477 |
| SRCNN | 24.134 | 0.590 | 26.062 | 0.616 | 23.548 | 0.547 | 20.285 | 0.443 | 19.719 | 0.542 |
| SRGAN | 22.957 | 0.555 | 25.731 | 0.598 | 22.638 | 0.518 | 20.954 | 0.438 | 18.288 | 0.488 |
| ESRGAN | 28.175 | 0.776 | 30.438 | <u>0.852</u> | 24.378 | 0.653 | 24.365 | 0.735 | 22.425 | 0.610 |
| USRGAN | 25.667 | 0.636 | 28.524 | 0.716 | <u>24.850</u> | 0.625 | 21.765 | 0.559 | 20.921 | 0.576 |
| SPSR | **30.397** | 0.772 | 30.397 | 0.844 | 24.457 | **0.6783** | <u>24.804</u> | 0.742 | 21.561 | 0.609 |
| CALGAN | <u>28.863</u> | **0.790** | **31.177** | 0.863 | **25.998** | <u>0.6778</u> | **25.290** | <u>0.763</u> | <u>23.182</u> | **0.625** |
| PSCAN (ours) | 28.544 | <u>0.776</u> | <u>30.784</u> | 0.743 | 24.678 | 0.668 | 24.377 | **0.773** | **23.201** | <u>0.617</u> |

From Fig. 3, it is evident that our proposed PSCAN demonstrates state-of-the-art performance in detail restoration and artifact removal. For instance, in the local region of the "Shoes" sample, the SR images produced by SRGAN, ESRGAN, and USRGAN fail to recover the clear white letters. In contrast, PSCAN, CALGAN, and SPSR show relatively better results, with PSCAN successfully

recovering clear white letters. The effectiveness of PSCAN in artifact removal is more noticeable in the "Building" sample: the SR images produced by the first three methods exhibit visible artifacts at the edges of the magnified fence, while the last three methods, including PSCAN, produce visually clearer edges, aligning with the quantitative PSNR and SSIM results shown in Table 1.

In Table 1, the best performance is highlighted in bold, and the second-best performance is underlined. Our proposed PSCAN achieves top-two performance in terms of PSNR and SSIM on five test sets and has a unique advantage in reducing artifacts near high-frequency details in super-resolution images. The performance variation of PSCAN across different test sets may be attributed to the diverse types of images in these datasets. Manga109 and Urban100, for example, are datasets of manga and urban scenes, respectively, which contain more high-frequency details compared to other datasets. This results in PSCAN's advantages being more pronounced in these two test sets.

### 5.2 More Challenging Root Image Super-Resolution

To further validate the advantages of the proposed PSCAN in high-frequency detail enhancement and artifact removal, we conducted comparative experiments on the "Root" dataset. Plant root images contain crucial information, and accurately obtaining their morphological parameters, such as root branch number and surface area, aids researchers in better identifying root traits, discerning plant growth and development, and studying the relationship between phenotypes and crop yield. Consequently, research on root image super-resolution is also of significant value. More importantly, due to the complex and intricate branching structure of roots, which manifests as rich edge texture details in images, this task presents a more challenging super-resolution problem, making it suitable for evaluating the performance of the proposed PSCAN method. This section compares the performance of PSCAN with classic super-resolution methods, including SRCNN, SRGAN, RCAN, USRGAN, SAN, CALGAN, and 3 traditional interpolation methods, from both visual effects and quantitative evaluation perspectives. Fig. 4 presents the SR image visualizations and the PSNR and SSIM comparison results on the "Root" dataset.

From Fig. 4, it can be observed that the SR images produced by interpolation methods are visually more blurred and lack high-frequency details, especially at the edges of the root structures. The nearest-neighbor interpolation method, in particular, exhibits severe checkerboard artifacts. SR images generated by deep learning-based methods, such as SRCNN, SRGAN, and USRGAN, show relatively better visual quality, with improved clarity at the edges of thicker primary roots compared to interpolation methods. However, these methods still suffer from significant artifact issues. The top three methods in terms of visual clarity are RCAN, CALGAN, and the proposed PSCAN. SR images based on RCAN are clear at the edges of thicker primary roots, but fail to recover finer side roots. CALGAN-based SR images recover the details of finer side roots but still exhibit artifacts. In contrast, PSCAN-based SR images perform best in recovering fine side roots, effectively replicating the complex side root structures seen in the HR images. In summary, the proposed PSCAN method achieves the best visual results, excelling in restoring fine texture details in the images.
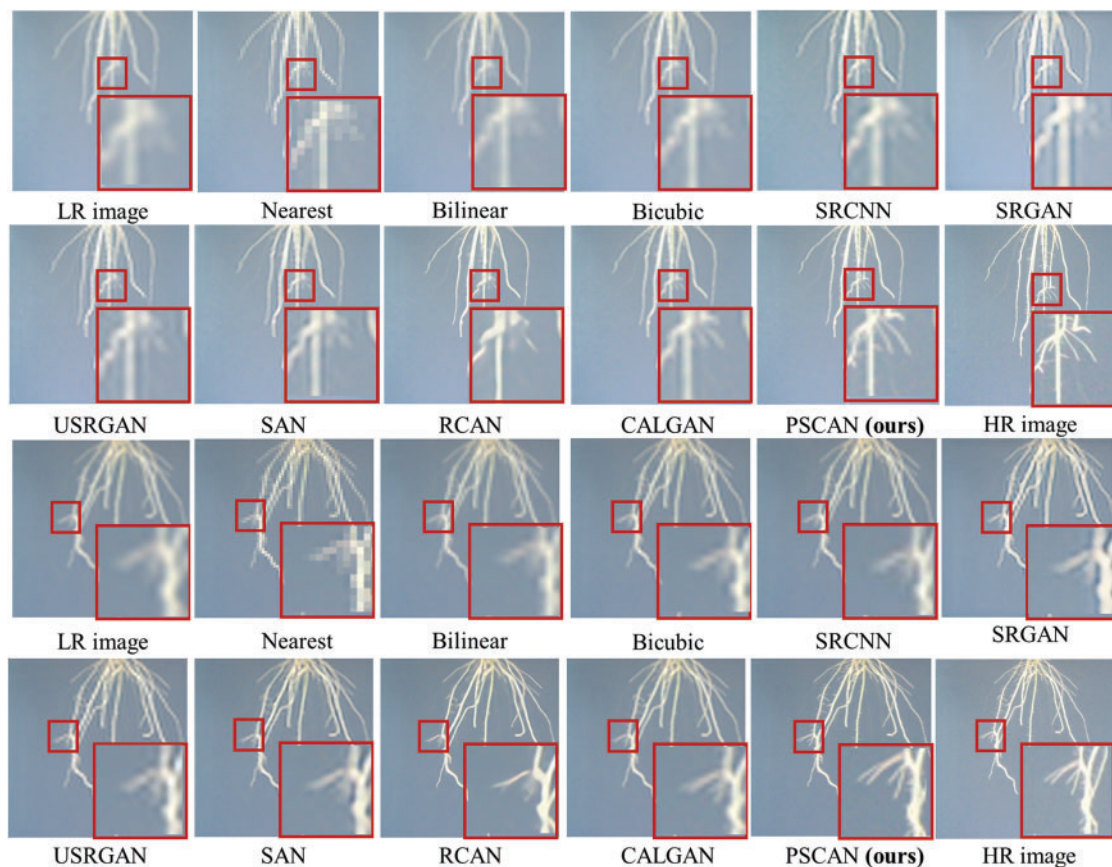
**Figure 4:** Results of 10 super-resolution methods on the "Root" dataset. There are four rows in total, with every two rows representing one set of results, thus showing two sample sets. Red boxes highlight and magnify local details in the super-resolution images

Table 2 presents the quantitative results of PSNR and SSIM for these 11 methods. Evidently, the proposed PSCAN method achieves state-of-the-art performance, outperforming the second-best RCAN by 0.84 in PSNR and 0.017 in SSIM.

**Table 2:** Results of several methods on "Root"

| Methods | Bicubic | SRCNN | SRGAN | USRGAN | SAN | RCAN | CALGAN | PSCAN |
|---|---|---|---|---|---|---|---|---|
| PSNR | 20.64 | 21.33 | 19.85 | 20.71 | 21.84 | 22.15 | 21.01 | **22.99** |
| SSIM | 0.6987 | 0.7102 | 0.7223 | 0.7115 | 0.7232 | 0.7354 | 0.7202 | **0.7525** |

### 5.3 Ablation Study

We conducted ablative experiments to demonstrate the efficacy of Pyramid Separable Channel Attention Blocks (PSCAB) and artifact loss in super-resolution tasks.

As shown in Fig. 5, compared to conventional convolutional structures, models utilizing PSCAB exhibit larger and more comprehensive high-response regions in feature heatmaps. This observation

underscores the positive impact of the pyramid separable convolutional structure in enlarging the model's receptive field.
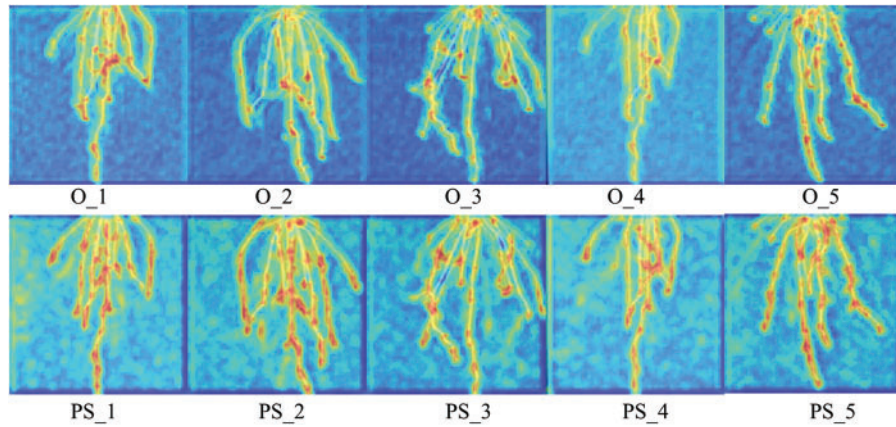


**Figure 5:** Ablation experiment results of Pyramid Separable Convolution Units. The figure is divided into two groups of subplots, each containing 5 samples. The upper group displays feature heatmaps using Pyramid Separable Channel Attention Blocks, while the lower group shows feature heatmaps using conventional convolutional channel attention blocks

Fig. 6 compares the results of the ablation study of the artifact loss factor and explains the mechanism of the artifact loss. It can be observed that the Gaussian high-pass filter extracts the edge texture details of the HR image.
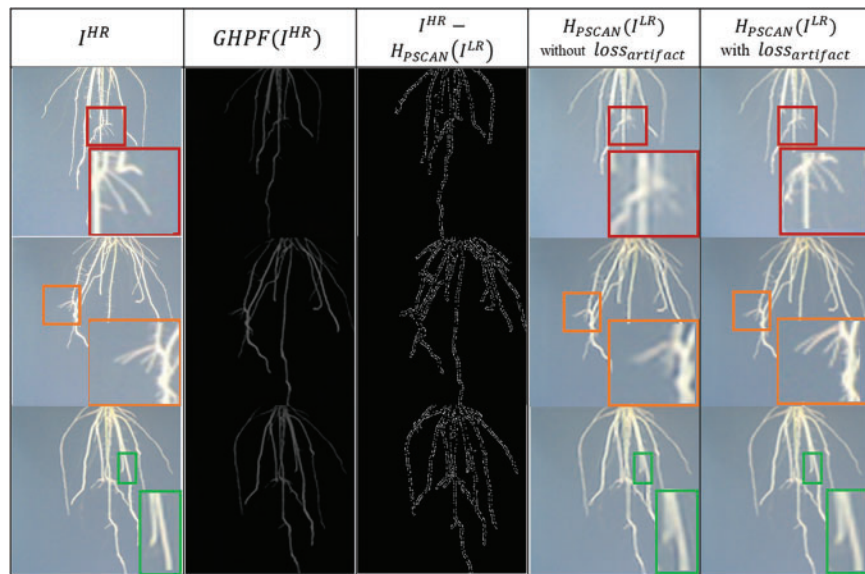


**Figure 6:** Visualization of the artifact loss ablation study results. From left to right: the original HR image, the edge information feature map of the HR image after Gaussian high-pass filtering, the residual feature map obtained by pixel-wise subtraction of the super-resolution image from the HR image, and the SR image before and after the ablation of the artifact loss factor. Color boxes highlight areas where differences are easily observed

The residual feature map contains both the high-frequency artifacts and edges of the root image, demonstrating the objective existence of artifacts in the SR image. The PSCAN model, by introducing the artifact loss factor, reduces the distribution difference between the residual information feature map and the edge detail feature map of the real image. As a result, the SR root image exhibits clear edge texture details, demonstrating the effectiveness of artifact loss in removing artifacts and preserving edge texture details in SISR task.

After 285 epochs, we report the PSNR and SSIM results on the root dataset, as detailed in Table 3. The pyramid separable convolution brings a 0.3 improvement in PSNR and a 0.011 improvement in SSIM to the model. The artifact loss factor brings a 0.64 improvement in PSNR and a 0.0168 improvement in SSIM to the model. From the results, it can be observed that the artifact loss contributes slightly more to the model's performance.

**Table 3:** Ablation study on root dataset

| Basic Model | With PSCU | With $loss_{\text{artifact}}$ | PSNR | SSIM |
|:---:|:---:|:---:|:---:|:---:|
| ✓ | | | 22.15 | 0.7354 |
| ✓ | ✓ | | 22.35 | 0.7357 |
| ✓ | | ✓ | 22.69 | 0.7415 |
| ✓ | ✓ | ✓ | **22.99** | **0.7525** |

## 6 Discussion and Conclusion

Existing SISR methods face challenges in reconstructing complex high-frequency details and effectively handling artifacts. To address these issues, this paper proposes the Pyramid Separable Channel Attention Network (PSCAN). Specifically, pyramid separable convolution units enable the model to fuse multi-scale features, and an artifact loss function is employed to remove artifacts and enhance high-frequency details in super-resolved images.

We conducted extensive training and testing on the Arabidopsis root image dataset and multiple public datasets. Results show that PSCAN achieves the best performance on challenging root image super-resolution tasks, with clear edges and state-of-the-art PSNR and SSIM scores, outperforming the second-best method by 0.84 and 0.017, respectively. It also excels in artifact removal on public datasets, demonstrating strong generalization capabilities.

Experimental results demonstrate that the proposed PSCAN method excels in preserving high-frequency edge details and eliminating artifacts in single-image super-resolution tasks. This makes it particularly valuable for challenging applications, such as plant root images and industrial crack images, underscoring its significant research and practical potential. This study still has some limitations, such as not achieving consistent high performance across different types of datasets. Addressing these challenges will be a key focus of our future research efforts.

**Author Contributions:** Congcong Ma: Conceptualization, Methodology, Validation, Formal analysis, Writing—original draft, Visualization. Jiaqi Mi: Conceptualization, Methodology, Software, Investigation, Writing—review & editing. Wanlin Gao: Resources. Sha Tao: Conceptualization, Methodology, Writing—review & editing. All authors reviewed the results and approved the final version of the manuscript.

**Availability of Data and Materials:** Not applicable.

**Ethics Approval:** Not applicable.

**Conflicts of Interest:** The authors declare that they have no conflicts of interest to report regarding the present study.

### References

[1] C. Saharia, J. Ho, W. Chan, T. Salimans, D. J. Fleet and M. Norouzi, "Image super-resolution via iterative refinement," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 4, pp. 4713–4726, Apr. 2023. doi: 10.1109/TPAMI.2022.3146844.

[2] Z. Wang, J. Chen, and S. C. Hoi, "Deep learning for image super-resolution: A survey," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 10, pp. 3365–3387, Oct. 2020. doi: 10.1109/TPAMI.2020.2982166.

[3] F. Deeba, S. Kun, F. A. Dharejo, and Y. Zhou, "Wavelet-based enhanced medical image super resolution," *IEEE Access*, vol. 8, pp. 37035–37044, Feb. 2020. doi: 10.1109/ACCESS.2020.2975292.

[4] M. Zhang, C. Zhang, Q. Zhang, J. Guo, X. Gao and J. Zhang, "ESSAformer: Efficient transformer for hyperspectral image super-resolution," presented at the IEEE/CVF Int. Conf. Comput. Vis., Paris, France, Oct. 2–6, 2023, pp. 23073–23084.

[5] R. Keys, "Cubic convolution interpolation for digital image processing," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 29, no. 6, pp. 1153–1160, Dec. 1981. doi: 10.1109/TASSP.1981.1163711.

[6] R. Nayak and D. Patra, "Enhanced iterative back-projection based super-resolution reconstruction of digital images," *Arab. J. Sci. Eng.*, vol. 43, no. 12, pp. 7521–7547, Dec. 2018. doi: 10.1007/s13369-018-3474-7.

[7] J. W. Zhao, Q. P. Yuan, J. Qin, X. P. Yang, and Z. H. Chen, "Single image super-resolution reconstruction using multiple dictionaries and improved iterative back-projection," *Optoelectron. Lett.*, vol. 15, no. 2, pp. 156–160, Apr. 2019. doi: 10.1007/s11664-018-06047-6.

[8] B. Lim, S. Son, H. Kim, S. Nah, and K. M. Lee, "Enhanced deep residual networks for single image super-resolution," presented at the IEEE Conf. Comput. Vis. Pattern Recognit. Workshops, Honolulu, HI, USA, Jul. 21–26, 2017, pp. 136–144.

[9] K. T. Gribbon and D. G. Bailey, "A novel approach to real-time bilinear interpolation," presented at the 2nd IEEE Int. Workshop Electron. Design, Test Appl., Perth, WA, Australia, Jan. 28–30, 2004, pp. 126–131.

[10] C. Liu, B. Yang, X. Zhang, and L. Pang, "IBPNet: A multi-resolution and multi-modal image fusion network via iterative back-projection," *Appl. Intell.*, vol. 52, no. 14, pp. 16185–16201, Dec. 2022. doi: 10.1007/s10489-022-03464-5.

[11] B. Xia, Y. Hang, Y. Tian, W. Yang, and J. Zhou, "Efficient non-local contrastive attention for image super-resolution," presented at the AAAI Conf. Artif. Intell., Vancouver, BC, Canada, Feb. 2–9, 2022, vol. 36, no. 3, pp. 2759–2767.

[12] P. Behjati, P. Rodriguez, C. Fernández, I. Hupont, A. Mehri and J. Gonzàlez, "Single image super-resolution based on directional variance attention network," *Pattern Recognit.*, vol. 133, Jan. 2023, Art. no. 108997. doi: 10.1016/j.patcog.2022.108997.

[13] K. C. Chan, X. Wang, X. Xu, J. Gu, and C. C. Loy, "GLEAN: Generative latent bank for large-factor image super-resolution," presented at the IEEE/CVF Conf. Comput. Vis. Pattern Recognit., Nashville, TN, USA, Jun. 19–25, 2021, pp. 14245–14254.

[14] W. S. Lai, J. B. Huang, N. Ahuja, and M. H. Yang, "Deep laplacian pyramid networks for fast and accurate super-resolution," presented at the IEEE Conf. Comput. Vis. Pattern Recognit., Honolulu, HI, USA, Jul. 21–26, 2017, pp. 624–632.

[15] R. Lan *et al.*, "Cascading and enhanced residual networks for accurate single-image super-resolution," *IEEE Trans. Cybern.*, vol. 51, no. 1, pp. 115–125, Jan. 2021. doi: 10.1109/TCYB.2020.2964601.

[16] J. Kim, J. K. Lee, and K. M. Lee, "Accurate image super-resolution using very deep convolutional networks," presented at the IEEE Conf. Comput. Vis. Pattern Recognit., Las Vegas, NV, USA, Jun. 27–30, 2016, pp. 1646–1654.

[17] C. Ledig *et al.*, "Photo-realistic single image super-resolution using a generative adversarial network," presented at the IEEE Conf. Comput. Vis. Pattern Recognit., Honolulu, HI, USA, Jul. 21–26, 2017, pp. 4681–4690.

[18] C. Dong, C. C. Loy, K. He, and X. Tang, "Image super-resolution using deep convolutional networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 2, pp. 295–307, Feb. 2016. doi: 10.1109/TPAMI.2015.2439281.

[19] Y. Zhang, Y. Tian, Y. Kong, B. Zhong, and Y. Fu, "Residual dense network for image super-resolution," presented at the IEEE Conf. Comput. Vis. Pattern Recognit., Salt Lake City, UT, USA, Jun. 18–22, 2018, pp. 2472–2481.

[20] Y. Guo *et al.*, "Closed-loop matters: Dual regression networks for single image super-resolution," presented at the IEEE/CVF Conf. Comput. Vis. Pattern Recognit., Seattle, WA, USA, Jun. 16–18, 2020, pp. 5407–5416.

[21] Y. Zhang, K. Li, K. Li, L. Wang, B. Zhong and Y. Fu, "Image super-resolution using very deep residual channel attention networks," presented at the ECCV, Munich, Germany, Sep. 8–14, 2018, pp. 286–301.

[22] B. Niu *et al.*, "Single image super-resolution via a holistic attention network," presented at the ECCV, Glasgow, UK, Aug. 23–28, 2020, pp. 191–207.

[23] T. Dai, J. Cai, Y. Zhang, S. T. Xia, and L. Zhang, "Second-order attention network for single image super-resolution," presented at the IEEE/CVF Conf. Comput. Vis. Pattern Recognit., Long Beach, CA, USA, Jun. 16–20, 2019, pp. 11065–11074.

[24] Goodfellow *et al.*, "Generative adversarial nets," *Adv. Neural Inf. Process. Syst.*, vol. 27, pp. 2672–2680, 2014.

[25] X. Wang *et al.*, "Esrgan: Enhanced super-resolution generative adversarial networks," presented at the ECCV Workshops, Munich, Germany, Sep. 8–14, 2018.

[26] K. Zhang, L. V. Gool, and R. Timofte, "Deep unfolding network for image super-resolution presented in IEEE/CVF Conf.," presented at the IEEE/CVF Conf. Comput. Vis. Pattern Recognit., Seattle, WA, USA, Jun. 16–18, 2020, pp. 3217–3226.

[27] C. Ma, Y. Rao, J. Lu, and J. Zhou, "Structure-preserving image super-resolution," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 11, pp. 7898–7911, Nov. 2021. doi: 10.1109/TPAMI.2021.3105586.

[28] J. Park, S. Son, and K. M. Lee, "Content-aware local gan for photo-realistic super-resolution," presented at the IEEE/CVF Int. Conf. Comput. Vis., Paris, France, Oct. 2–6, 2023, pp. 10585–10594.

[29] H. Chen *et al.*, "Pre-trained image processing transformer," presented at the IEEE/CVF Conf. Comput. Vis. Pattern Recognit., Nashville, TN, USA, Jun. 19–25, 2021, pp. 12299–12310.

[30] J. Liang, J. Cao, G. Sun, K. Zhang, and R. Timofte, "SwinIR: Image restoration using swin transformer," presented at the IEEE/CVF Int. Conf. Comput. Vis., Montreal, QC, Canada, Oct. 11–17, 2021, pp. 1833–1844.

[31] M. Bevilacqua, A. Roumy, C. Guillemot, and M. L. Alberi-Morel, "Low-complexity single-image super-resolution based on nonnegative neighbor embedding," presented at the BMVC, Surrey, UK, Sep. 3–7, 2012.

[32] R. Zeyde, M. Elad, and M. Protter, "On single image scale-up using sparse-representations," in *Curves Surfaces*, Avignon, France: Springer, Jun. 24–30, 2012, pp. 711–730. doi: 10.1007/978-3-642-27413-8_47.

[33] J. B. Huang, A. Singh, and N. Ahuja, "Single image super-resolution from transformed self-exemplars," presented at the IEEE Conf. Comput. Vis. Pattern Recognit., Boston, MA, USA, Jun. 7–12, 2015, pp. 5197–5206.

[34] Y. Matsui *et al.*, "Sketch-based manga retrieval using manga109 dataset," *Multimed. Tools Appl.*, vol. 76, no. 20, pp. 21811–21838, Sep. 2017. doi: 10.1007/s11042-016-4020-z.

[35] J. F. Ruiz-Munoz, J. K. Nimmagadda, T. G. Dowd, J. E. Baciak, and A. Zare, "Super resolution for root imaging," *Appl. Plant Sci.*, vol. 8, no. 7, Jul. 2020, Art. no. e11374. doi: 10.1002/aps3.11374.

[36] J. Mi *et al.*, "A method of plant root image restoration based on GAN," *IFAC-PapersOnLine*, vol. 52, no. 30, pp. 219–224, Dec. 2019. doi: 10.1016/j.ifacol.2019.12.410.

[37] D. Mishra, S. Chemweno, O. Hadar, O. Ben-Tovim, N. Lazarovitch and J. E. Ephrath, "Unsupervised image super-resolution for root hair enhancement and improved root traits measurements," *IEEE Trans. AgriFood Electron.*, vol. 2, no. 1, 2024. doi: 10.1109/TAFE.2024.3359660.