Engineering Applications of Artificial Intelligence 139 (2025) 109535

Contents lists available at ScienceDirect

Engineering Applications of Artificial Intelligence

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

Dual-path aggregation transformer network for super-resolution with images occlusions and variability

Qinghui Chen [a,e], Lunqian Wang [b], Zekai Zhang [a], Xinghua Wang [b], Weilin Liu [b], Bo Xia [b], Hao Ding [b], Jinglin Zhang [a,c,*], Sen Xu [d], Xin Wang [b]

[a] School of Control Science and Engineering, Shandong University, Jinan, 250061, China
[b] School of Information Science and Engineering, Linyi University, Linyi, 276000, China
[c] Shandong Research Institute of Industrial Technology, Jinan, 250061, China
[d] Information Science and Engineering, YanCheng Institute of Technology, Yancheng, 224000, China
[e] Laoshan Laboratory, Qingdao, 266237, China

ARTICLE INFO

Keywords:
Super-resolution
Transformer
Satellite cloud images

ABSTRACT

While Transformer-based approaches have recently achieved notable success in super-resolution, their extensive computational requirements impede widespread practical adoption. High-resolution meteorological satellite cloud imagery is essential for weather analysis and forecasting. Enhancing image resolution through super-resolution techniques facilitates the accurate identification and localization of geographic features by meteorological systems. However, current super-resolution methods fail to restore the intricacies of cloud formations and complex regions fully. This research introduces a novel dual-path aggregation Transformer network (DPAT) tailored to enhance the super-resolution of meteorological satellite cloud images. The DPAT network adeptly captures cloud imagery's subtle details and textures, effectively addressing occlusions and the variability inherent in satellite imagery. It bolsters the model's ability to manage the complex attributes of cloud images through the introduction of the Dual-path Aggregation Self-Attention (DASA) mechanism and the Multi-scale Feature Aggregation Block (MFAB), thereby enhancing performance in processing intricate cloud features. The DASA mechanism synthesizes features across spatial, depth, and channel dimensions via a dual-path approach, thoroughly exploiting feature correlations. The MFAB, designed to supplant the multilayer perceptron, incorporates shift convolution and a multi-scale interaction block to augment feature information, compensating for the deficiency in local information absorption due to fixed receptive fields. Experimental outcomes indicate that DPAT delivers superior super-resolution outcomes. With a parameter count of only 32% of the Enhanced Deep Residual Network (EDSR) or 77% of the Image Restoration using Shift Window Transformer (SwinIR), DPAT matches SwinIR's performance on the satellite cloud dataset. Moreover, DPAT balances accuracy and parameter economy across various datasets. This technology is expected to improve image super-resolution capabilities in multiple fields such as human action recognition and industrial recognition, and indirectly improve the accuracy of image perception tasks.

1. Introduction

Deep learning has developed rapidly in recent years, and neural network-based SR methods have emerged. CNN-based SR models are capable of end-to-end learning with large amounts of training data and can overcome some drawbacks found in traditional reconstruction algorithms (Keshk and Yin, 2017). The pioneering work SRCNN (Chao et al., 2016) first uses a convolutional neural network (CNN) to learn the mapping from LR inputs to HR targets and dramatically outperforms the traditional single-image super-resolution (SISR) approach. Following EDSR (Lim et al., 2017), a series of well-designed CNN architectures (Muhammad et al., 2018; Wang et al., 2018; Ying et al., 2017; Yulun et al., 2018) and visual attention mechanisms (Ben et al., 2020; Mei et al., 2021; Tao et al., 2019) were introduced to improve the performance of CNN-based SR models. There are also some models that combine traditional methods with deep learning methods (Li et al.,

* Corresponding author. School of Control Science and Engineering, Shandong University, Jinan, 250061, China.
E-mail addresses: 202420785@mail.sdu.edu.cn (Q. Chen), 210854002032@lyu.edu.cn (L. Wang), 202420810@mail.sdu.edu.cn (Z. Zhang), 210854002052@lyu.edu.cn (X. Wang), 210854002027@lyu.edu.cn (W. Liu), jinglin.zhang@sdu.edu.cn (J. Zhang), xusen@ycit.cn (S. Xu), wangxin@lyu.edu.cn (X. Wang).

https://doi.org/10.1016/j.engappai.2024.109535
Received 7 June 2024; Received in revised form 20 September 2024; Accepted 21 October 2024
Available online 31 October 2024
0952-1976/© 2024 Elsevier Ltd. All rights are reserved, including those for text and data mining, AI training, and similar technologies.

2020; Xie et al., 2018). Self-attention-based network structures, especially SwinIR (Jingyun et al., 2021), can utilize global information from shallow to deep layers more effectively, further enhancing the quality of SR. The SR effect of satellite cloud images has also been greatly improved. However, the features in satellite cloud images have strong irregularities, which seriously impact the effectiveness of SR. Common channel or spatial attention mechanisms are difficult to adapt to the complex features in satellite cloud images, resulting in CNN-based models that can not fully exploit their performance. Self-attention-based networks have long computing time and excessive resource consumption, while window self-attention tends to ignore local information. In addition, by local attribution maps (LAM) (Gu and Dong, 2021), an attribution analysis method, C. Xiangyu et al. (Chen et al., 2023) found that SwinIR does not exploit more pixels than the CNN-based methods in SR tasks. Although SwinIR obtains higher quantitative performance on average, the local information utilized by SwinIR is limited.

Satellite cloud images consist of images of cloud cover and ground surface features on Earth as observed by meteorological satellites, including infrared cloud images, visible cloud images, and water vapor images. Satellite cloud images can assist people in identifying different weather systems and provide a basis for weather analysis and forecasting. Satellite cloud images can also contribute to building weather information in areas that lack weather observation stations, such as oceans, deserts, and plateaus. Resolution is an important indicator of the quality of satellite cloud images. High-resolution (HR, see Table 1) satellite cloud images can facilitate experts to analyze geographic information more accurately. However, the spatial resolution of satellite cloud images has not been able to meet the growing demand for various weather applications (B.G. Henderson et al., 2005). Utilizing algorithms to enhance resolution proves to be a more efficient strategy, as technological capabilities and associated expenses frequently restrict advancements in hardware devices. Image super-resolution (SR) is the process of recovering a HR image from a low-resolution (LR) image or sequence of images. HR means that the image has a high pixel density and more details that are essential in many practical applications. SR technology has been successfully applied in various fields such as medicine (K. Umehara et al., 2018; X. Zhao et al., 2019), remote sensing (C. Tuna et al., 2018; W. Xu et al., 2018), and energy engineering (Y. Da Wang et al., 2019). Due to the limitations of satellite imaging principles and interference during the reception process, the spatial resolution of traditional satellite cloud images often cannot meet the needs of refined monitoring, especially since the resolution of infrared cloud images is often low, and there is a problem of unclear grey levels. The multi-scale characteristics of satellite cloud images and the chaotic effects of the atmosphere make the image sequence change law complex, which brings difficulties to the comprehensive use of multi-channel cloud image information. Currently, SR methods are broadly classified into three categories: interpolation methods (Y. Kawai et al., 2006; W.H. Kim and Kim, 2010), reconstruction methods (A. Garg et al., 2014; H. Demirel and Anbarjafari, 2011; R.B. Ahire et al., 2013) and learning methods (Z. Ying et al., 2016; R. Zeyde et al., 2012; W. Jin et al., 2014).

Table 1
Main abbreviations used in this work and their full names.

Abbreviations	Full names
DPAT	Dual-path Aggregation Transformer Network
DASA	Dual-path Aggregation Self-Attention
MFAB	Multi-scale Feature Aggregation Block
EDSR	Enhanced Deep Residual Network
HR	High-resolution
SR	Super-resolution
LR	Low-resolution
NCCI	Natural Color Cloud Image
ViT	Vision Transformer
MTB	Multi-path Transformer Block
SF	Shallow Feature Extraction Module
DF	Deep Feature Extraction Module

The interpolation-based method effectively exploits the correlation of neighboring pixels in the original image to accomplish the SR task. Kim et al. (W.H. Kim and Kim, 2010) reduced the artifacts of SR results of cloud images with the assistance of estimated errors and interpolation algorithms. However, the interpolation method tends to lose high-frequency information and leads to blurring of the generated images. The reconstruction-based method models the acquisition process of LR images to construct a priori constraints for HR images and uses them to accomplish the SR task. A. Garg et al. (A. Garg et al., 2014) improved the resolution of satellite images by introducing discrete wavelet transform (DWT), and they interpolated different sub-band satellite images after DWT decomposition and then combined the interpolated images to complete the SR task. However, the reconstruction method only uses the information from the LR image, and the high-frequency information utilized for reconstruction is insufficient to satisfy the mission requirements. Learning-based methods obtain a priori knowledge by learning from training samples, and traditional learning methods mainly refer to dictionary methods. LR dictionaries extract overlapping patches from images to encode them, and then pass sparse coefficients to HR dictionaries to achieve reconstruction (J. Yang et al., 2010). All these methods do not utilize the information of LR images completely enough, resulting in an insufficient representation of the image contour textures.

In this work, to address these issues, authors design a dual-path aggregation self-attention (DASA) based on the striped window and utilize channel attention to activate the performance of self-attention. DASA can fully explore the correlations between features in the three dimensions of space, channel, and depth with fewer parameters and computational cost. Because the operation mode of MLP is not suitable for spatial information-rich data like images, authors also redesign a novel multi-scale feature aggregation block (MFAB) to replace MLP to further enhance the SR effect of satellite cloud images. Then authors propose a dual-path aggregation Transformer network (DPAT) based on DASA and MFAB. The quantitative trade-off between the performance and complexity of the models is shown in Fig. 1, which demonstrates the superiority of the proposed network. DPAT effectively fills the gap of existing methods through a dual-path mechanism. Using self-attention in two dimensions, space and channel, feature information is retained through one-dimensional convolution, and the other path extracts deeper features through a convolution group with channel attention. Feature aggregation of different dimensions is achieved, which enhances the model's ability to capture global context. The design of DPAT considers the multi-scale characteristics of satellite cloud images and the chaos of the atmosphere. The multi-scale features are refined through the fusion network so that DPAT can effectively reconstruct high-frequency information such as edges and textures. Contributions can be summarized as follows.

1. **Improved and Efficient Self-Attention:** The paper introduces a novel DASA that explores feature correlations across spatial, channel, and depth dimensions more effectively and with fewer parameters and computational costs than traditional methods.
2. **Novel and Precise Transformer Layer:** A new multi-scale Transformer layer (MTL) is proposed, featuring a MFAB designed to replace the traditional MLP, enhancing the SR effect by better associating features from different receptive fields.
3. **Lightweight and Fast Network:** The proposed DPAT is a streamlined model, using 32% of EDSR or 77% of SwinIR's parameters. It matches SwinIR's accuracy on satellite imagery and excels on the NCCI (Jinglin Zhang et al., 2021) dataset for super-resolution. DPAT also performs well on other benchmarks, like the DIV2K (Bee Lim et al., 2017) dataset, showcasing its effectiveness in capturing high-frequency information such as edges and textures.

This work addresses the gap in existing SR methods by introducing a

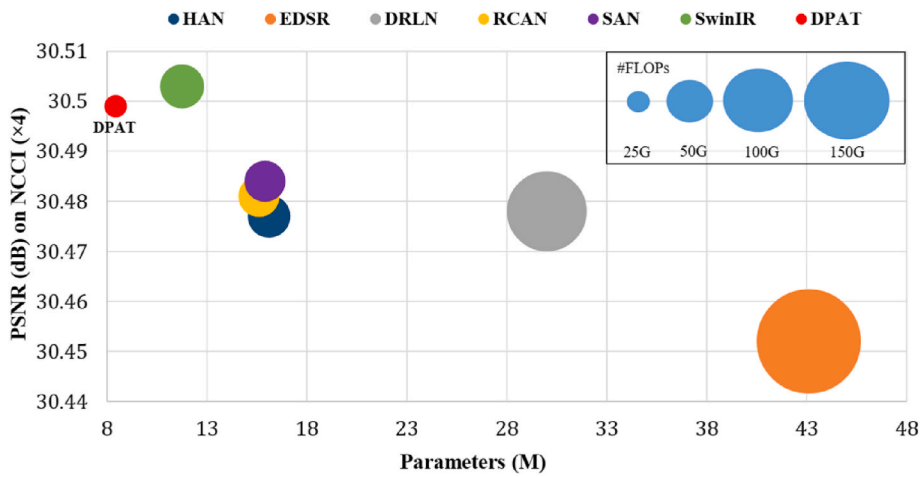

Fig. 1. Qualitative trade-off comparison between model performance and complexity of × 4 image SR on NCCI dataset. The comparison results demonstrate the superiority of DPAT.

dual-path mechanism that retains feature information through one-dimensional convolution while extracting deeper features through a convolution group with channel attention. The contributions of this research not only advance the field of image SR but also hold significant implications for meteorological applications and the enhancement of satellite cloud image analysis. The rest of the content is organized as follows. Section 2 briefly reviews the related works of SR Frameworks, Artificial Intelligence Progress and Transformer-based Frameworks. Section 3 comprehensively describes the proposed method. Section 4 presents extensive experimental results and visual comparisons. Finally, Section 5 summarizes the effectiveness and prospects of the proposed method.

2. Related work

2.1. Super-resolution frameworks

Rapidly evolving techniques in deep learning have allowed SR methods to be actively explored and achieve state-of-the-art performance. CNN can accurately extract local features of images and ensure that the features are translation invariant. Since SRCNN introduced the CNN into the SR task, SR models have been updated iteratively by gradually adding convolutional and activation layers.

However, the excessive number of parameters and quantity of computation will reduce the efficiency. Then people introduce attention mechanisms to enhance SR effectiveness. RCAN (Yulun Z et al., 2018) introduced channel attention to enhance the importance of features that benefit the SR task, but channel attention has difficulties in preserving information texture and recovering natural details. HAN (Ben N et al., 2020) combined channel spatial attention and layer attention on the basis of RCAN to make the network more focused on the feature channels with rich information. NLSA (Yiqun Mei et al., 2021) brought significant improvements to the SR task through non-local attention (NLA), which is a non-local attention mechanism that exploits the intrinsic feature correlation in natural images. Inspired by vision Transformers (Ze L et al., 2021), self-attention has been employed in SR to capture long-term adaptability. SwinIR (Jingyun L et al., 2021) made the SR effect substantially improved with the assistance of the Transformer. However, both CNN-based and Transformer-based models have their own limitations in the SR task of satellite cloud images. ST Seydi et al. (ST Seydi et al., 2024) introduced a novel framework for detecting changes using bi-temporal hyperspectral images, built upon a dual-stream deep feature extraction process. However, the framework may be limited by the computational complexity. Jiang et al. (2013) proposed a SR method that simultaneously improves image resolution and visual quality by proposing a practical solution combining manifold regularization with sparse support regression. They also introduced graph embedding SR, which can super-resolve HR images from a single LR observation. However, the method could face challenges in handling diverse image degradations. Lepcha et al. (Lepcha et al., 2022) introduced a piecewise linear regression strategy to learn image-specific LR to HR mappings. Hadamard patterns generated from LT training patches were used to classify LR-HR training patch pairs into different classes. Finally, a piecewise linear regression strategy was adopted to learn each class's mapping relationship between LR and HR spaces. Despite its innovative approach, the strategy might be constrained by the quality of the training patches and the assumption of piecewise linearity in image mappings. Chan et al. (2021) introduced four basic functions (such as alignment, propagation, aggregation, and upsampling) to untie the knot and reprocess several very important components of video SR. The strategy uses a methodical approach to describe how such gains can be achieved and explore potential difficulties. However, the effectiveness of the strategy might be limited by the complexity of the video content and the need for precise alignment in the initial stages. Yang et al. (2021) introduced an enhanced Dirac residual block based on the Dirac convolution model, which uses trainable parameters to control the balance between convolution and jump connections adaptively and implements global jump connections to enhance the nonlinear mapping capability of the model. The model shows excellent performance both visually and quantitatively, but could be sensitive to the choice of hyperparameters.

2.2. Artificial Intelligence Progress and Transformer-based frameworks

Artificial intelligence technology has made important progress. Akkem Y et al. (Akkem Y et al., 2023) reviewed the application of artificial intelligence technology in agriculture, using machine learning and deep learning to make appropriate crop recommendations to reduce future food shortages. J Agnihotri et al.(J Agnihotri et al., 2023) understood this conflicting phenomenon by modeling the Mississippi River basin using the Noah land surface model with the multiphysics option and a routing application for parallel computation of flow. MA Farmani et al.(MA Farmani et al., 2024) improve the physical representation of soil hydrology in the widely used Noah-MP land surface models and provide alternative soil hydrology/hydraulic schemes. MJ Ebadi et al. (MJ Ebadi et al., 2021) consider the color or luminance variations of a spatial position in a series of frames as input data points in Euclidean space R or R3. The video data points are approximated using progressive and iterative approximation for least square fitting. K Safari et al. (K Safari et al., 2022) solved the complex design problem of a posture

adjustment controller by introducing an intelligent controller based on brain emotion learning and adjusted by a fuzzy inference system. These studies demonstrate the contribution of artificial intelligence technology in various fields.

Alexey D et al. (Alexey D et al., 2021) first introduced the Transformer structure to vision tasks and designed the ViT. ViT exhibits more efficient performance than CNN-based models. So Ze L et al. introduced some ideas of induction bias in CNN into Transformer-based architecture and proposed the shifted window (Swin) based Transformer. Swin partitions the input features and applies self-attention to each partition separately, which reduces the quantity of computation and makes Transformer-based structures more practical for vision tasks. The Transformer-based models have since been iterated and have achieved outstanding results in many areas of computer vision. Jingyun L et al. (Jingyun L et al., 2021) simplified Swin for SR and achieved excellent results. HAT (Xiangyu Chen et al., 2023) combined both channel attention and window-based self-attention schemes, thus making use of their complementary advantages of being able to utilize global statistics and strong local fitting capability. Jinpeng S et al. (Jinpeng Shi et al., 2023) proposed the efficient striped window Transformer (ESWT) in order to balance the performance and complexity of the Transformer model.

Transformer is able to better capture contextual information by learning the dependencies between different positions in the sequence through a self-attention mechanism. However, Transformer lacks modeling of local information, which is the opposite of CNN. Authors combine convolution and self-attention to further exploit the potential of the Transformer and apply it to the SR task of satellite cloud image.

Authors explore the potential of incorporating physical priors or domain-specific knowledge into Transformer-based super-resolution frameworks to enhance the fidelity of reconstructed HR satellite cloud images (Sun, Y. et al., 2024). The integration of physical models of satellite imaging, such as atmospheric scattering models and sensor response functions, can more accurately simulate the image degradation process (Kabir, S. et al., 2020). Additionally, leveraging domain knowledge from meteorology and remote sensing can guide the design of super-resolution algorithms to better handle satellite cloud image characteristics. The development of adaptive feature extraction algorithms that reflect the physical properties of cloud layers and surface features can improve the quality and accuracy of reconstructed images (Li Y. et al., 2022). The fusion of multimodal data, such as multispectral, infrared, and visible light images, can enhance the effectiveness of super-resolution reconstruction by utilizing information from different sensors. Combining deep learning techniques with physical models can learn and simulate physical processes while incorporating physical model constraints as regularization terms in the learning process to improve the model's generalization capability and reconstruction quality (Wi, S. et al., 2022). Optimizing attention mechanisms to capture the features of satellite cloud images better and designing loss functions that measure the quality and physical consistency of reconstructed images can also guide the optimization direction during model training (Xu, M. et al., 2022).

3. Proposed method

3.1. The Overall structure

As illustrated in Fig. 2, the proposed DPAT is composed of three parts: the shallow feature extraction module (SF), the deep feature extraction module (DF), and the image reconstruction module (IR). DPAT is designed for SR tasks, particularly for enhancing the LR satellite cloud images. SF extracts primitive features from the LR satellite cloud image through a single convolution operation. DF consists of multiple cascading MTBs and a 3x3 convolution layer at the end, which is responsible for extracting deep features used to reconstruct the HR image. IR takes the deep features and, with the addition of shallow features through a long skip connection, reconstructs the HR image using a pixel-shuffle layer and a reduced-dimension convolution layer. As a key component of the MTL, DASA uses a striped window-based self-attention mechanism to establish remote dependencies efficiently. It employs dual-path convolution and channel attention to capture feature correlations in spatial, channel, and depth dimensions. MFAB is designed to compensate for the limitations of attention mechanisms by using shift convolution and multi-scale interactive operations to capture and aggregate spatial information and features from different scales. The model's innovative approach lies in its dual-path self-attention mechanism and multi-scale feature aggregation, which help prevent the loss of important features such as water clouds and effectively handle the unique challenges of satellite cloud images.

The LR satellite cloud image is input into the network, and the primitive feature F_l is extracted by the SF. This part consists of a single 3×3 convolution, termed as $H_{SF}(\cdot)$:

$$F_l = H_{SF}(I_{LR}) \tag{1}$$

where I_{LR} represents the LR satellite cloud image. The F_l is then sent to DF.

DF consists of m cascading MTBs and utilizes a 3×3 convolution layer as the tail. The whole process of DF termed as $H_{DF}(\cdot)$:

$$F_h = H_{DF}(F_l) = H_C\left(H_{MTB}^m\left(H_{MTB}^{m-1}\left(\cdots H_{MTB}^1(F_l)\cdots\right)\right)\right) \tag{2}$$

where F_h denotes the deep features utilized to reconstruct the HR image. $H_{DF}(\cdot)$ denotes the DF. $H_C(\cdot)$ denotes the 3×3 convolution of the tail of DF. $H_{MTB}^i(\cdot)(i \in 1, 2, \cdots, m)$ denote the $i-th$ MTB. In addition, each MTB

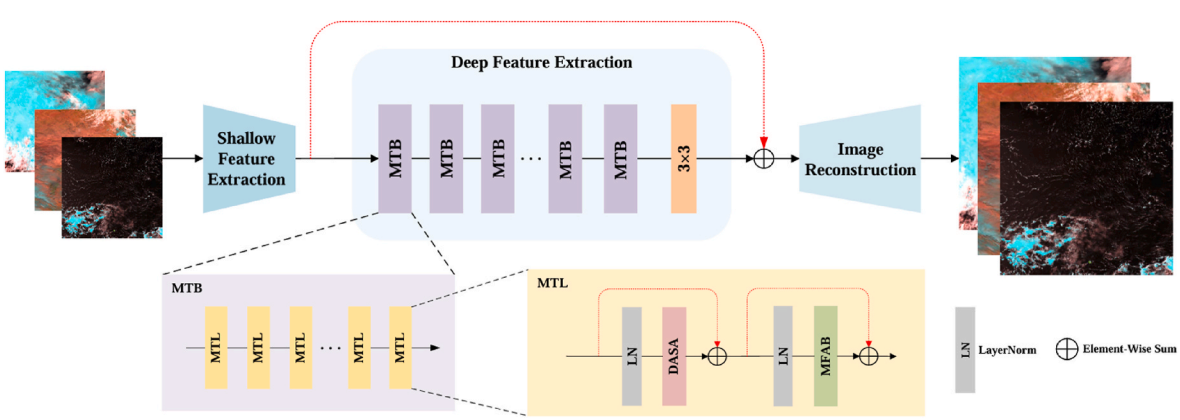

Fig. 2. Overview of DPAT. The deep feature extraction module is composed of m cascading MTBs. Each MTB consists of n cascading MTLs. Each MTL consists of LN, DASA, and MFAB. Pixel-shuffle is adopted for the up-sampling in the image reconstruction module.

contains n cascading dual-path Transformer layers (MTLs). The function $H^i_{MTB}(\cdot)$ can be further expressed as follows:

$$F_i = H^i_{MTB}(F_{i-1}) = H^n_{MTL}\left(H^{n-1}_{MTL}\left(\cdots H^1_{MTL}(F_{i-1})\cdots\right)\right) \tag{3}$$

where F_i denotes the output of H^i_{MTB}. $H^j_{MTL}(\cdot)(j \in 1, 2, \cdots, n)$ denote the $i-th$ MTL.

Finally, F_h is sent to the IR, and the shallow feature F_l is added to the reconstruction module by a long skip connection. The IR restores the image through $H_{IR}(\cdot)$:

$$I_{SR} = H_{IR}(F_h + F_l) \tag{4}$$

where I_{SR} represents the reconstructed HR image. $H_{IR}(\cdot)$ contains a pixel-shuffle layer and a reduced-dimension convolution layer. The following sections will detail the structure and function of each module in the network.

3.2. Multi-path Transformer layer (MTL)

Many works illustrate that convolution can facilitate the Transformer to get a better visual representation or achieve easier optimization (L. Kunchang et al., 2023; Haiping Wu et al., 2021a; Tete Xiao et al., 2021). Therefore authors combine CNN and Transformer in a new way to accomplish SR tasks. As shown in Fig. 2, the base module of DPAT is MTL, which mainly consists of DASA and MFAB. Unlike the original Transformer and SwinIR, feature extraction is basically accomplished by convolution operation in MTL. DASA employs the striped window-based self-attention mechanism, which enables the establishment of remote dependencies more efficiently. Both GRL (Yawei Li et al., 2023) and ESWT (Jinpeng Shi et al., 2023) have demonstrated that the striped window mechanism can capture contextual information from different dimensions and thus can better model long-term dependencies, as shown in Fig. 4. In addition, to better aggregate multi-scale information, authors replace the MLP in the original Transformer layer with the MFAB. Given an input feature F_x, the whole process of MTL can be expressed as:

$$\begin{aligned} F_s &= H_{DASA}\left(H_{Laynorm}(F_x)\right) + F_x \\ F_{out} &= H_{MFAB}\left(H_{Laynorm}(F_s)\right) + F_s \end{aligned} \tag{5}$$

where F_s denotes the output of DASA, $H_{DASA}(\cdot)$ denotes DASA, $H_{Laynorm}(\cdot)$ denotes LayerNorm (LN) operation, $H_{MFAB}(\cdot)$ denotes MFAB, and F_{out} is the final output.

3.3. Dual-path aggregation self-attention

DASA based on the striped window self-attention mechanism is a critical part of MTL. As shown in Fig. 5, even with the assistance of residual connection, water cloud features in the deeper layers of the SR network are severely lost, which is detrimental to the reconstruction of HR images. The dual-path approach is applied to prevent the loss of water cloud features. Specifically, one path retains feature information through 1D convolution, and the other path extracts deeper features through the convolution group with channel attention. In this way, features of different depths can be obtained to avoid the loss of features such as water clouds, thus reducing the error in HR image reconstruction.

For the input feature with shape $C \times H \times W$, the original Transformer layer needs to perform a series of time-consuming operations to reshape its shape to $HW \times C$ first, making the convolution operation impossible to apply. However, many studies have shown that convolution can facilitate the Transformer to obtain better visual representation or achieve easier optimization (Haiping Wu et al., 2021; Tete Xiao et al., 2021a; Kun Yuan et al., 2021). Furthermore, HAT has shown that channel attention can benefit self-attention to activate more pixels. Therefore, in DASA, features are first processed by dual-path convolution before computing self-attention. The dual-path features are then split by striped windows of the same size and computed as different *queries* by linear mapping. Different *queries* are assigned weights and aggregated to obtain a *query* for self-attention computation, which contains complete and rich feature information. The *value* is obtained by partitioning and reshaping the features after aggregation so that it is guaranteed to correspond to the aggregated *query*. Finally, the *query*, *key*, and *value* are processed according to the self-attention formula. As shown in Fig. 3. DASA relies on the dual-path structure to more fully fuse the information of different frequencies and avoid the loss of low-frequency features. And the designed self-attention mechanism can achieve richer information interactions with fewer parameters. Specifically, the input feature F_x generates F_m and F_n after dual-path convolution. The processing can be expressed as follows:

$$\begin{aligned} & F_x = C_0(F_{in}), \\ & F_m = C_1(F_x), \\ F_n = H_{CAB}(F_x) \quad & = C_4(H_{CA}(C_3(GELU(C_2(F_x))))) \end{aligned} \tag{6}$$

where $C_1(\cdot)$ and $C_4(\cdot)$ denotes 1×1 convolution. $H_{CAB}(\cdot)$ denotes attentional convolution block. $C_0(\cdot)$, $C_2(\cdot)$, and $C_3(\cdot)$ denote 3×3 convolution. $GELU(\cdot)$ denotes the Gaussian error linear unit (GELU) activation function. $H_{CA}(\cdot)$ is channel attention. $F_m \in R^{B\times C\times H\times W}$ and $F_n \in R^{B\times C\times H\times W}$ will be partitioned according to the window size provided. Since the network adopts the striped window, two parameters (w_1, w_2) will be provided to specify the window size. F_m and F_n after partitioning are calculated by linear mapping separately to obtain $Q_m \in R^{\frac{BMHW}{w_1 w_2}\times w_1 w_2 \times C}$ and $Q_n \in R^{\frac{BMHW}{w_1 w_2}\times w_1 w_2 \times C}$. M denotes the number of *heads* in the self-attention. Then the two *queries* are given weights (α, β) and aggregated by matrix addition.

$$Q = \alpha Q_m + \beta Q_n \tag{7}$$

F_m and F_n are combined, and calculated by linear mapping to obtain the V. Q, V are used for attention calculation, and self-attention is expressed as:

$$F_{sa} = Softmax\left(\frac{QQ^T}{\sqrt{d}}\right)V \tag{8}$$

where $Softmax(\cdot)$ denotes the Softmax function, d denotes the *query/key* dimension, Q^T denotes the *key*, and F_{sa} denotes the output features of DASA. F_{sa} will be converted to the shape $B \times C \times H \times W$.

DASA utilizes dual-path convolution and channel attention to promote self-attention to explore feature correlations in three dimensions: space, depth, and channel.

In addition, DASA establishes remote dependencies faster with the striped window.

The striped window self-attention mechanism focuses on local spatial features, and channel self-attention is performed along the channel dimension. Through the collaborative work of 1D convolution and channel attention mechanism, the design of DASA allows the network to capture global and local features without requiring too many computing resources. It can quickly capture and reconstruct the features required for HR images and reduce complexity.

DASA utilizes a self-attention mechanism based on striped windows to effectively establish long-range dependencies. As an important feature selection component, MTL adopts dual-path convolution and channel attention to capture spatial, channel, and depth dimensions feature correlations. MFAB compensates for the limitations of the attention mechanism by using shifted convolution and multi-scale interaction operations. It utilizes rich regional features, allows them to interact spatially with the target area, and converges adaptively to alleviate the adverse effects of satellite cloud image blur and cloud occlusion on SR tasks.

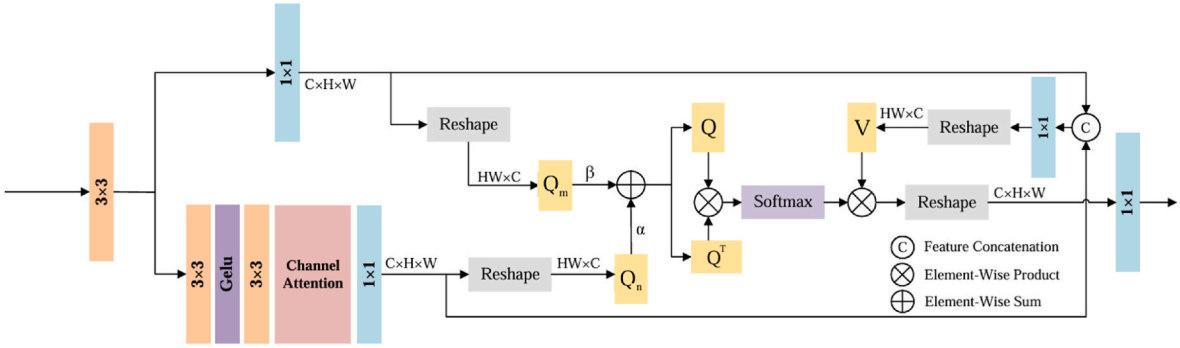

Fig. 3. Illustration of DASA, which utilizes dual-path convolution and channel attention to further stimulate self-attention. And the self-attention is based on the striped window. DASA can fully explore the correlations between features in the three dimensions of space, channel, and depth with fewer parameters and computational cost.

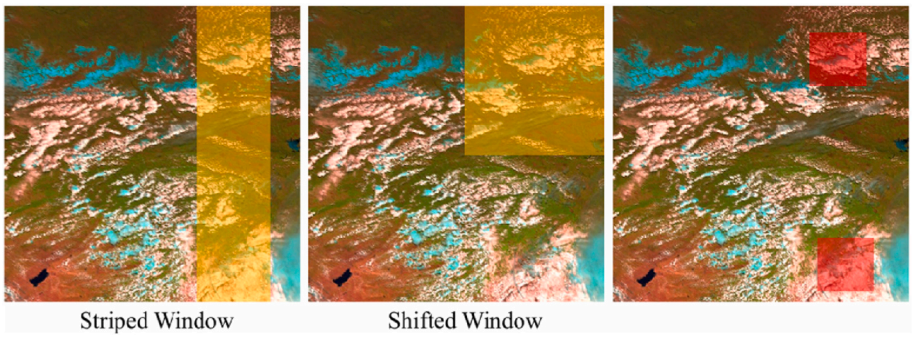

Fig. 4. Comparison of striped window and shifted window. The yellow area in the image is where the window covers, and the red area is where the association is assumed to be established. It can be seen that the striped window is able to establish remote dependencies well, but the shifted window needs to be shifted to catch the associated features.

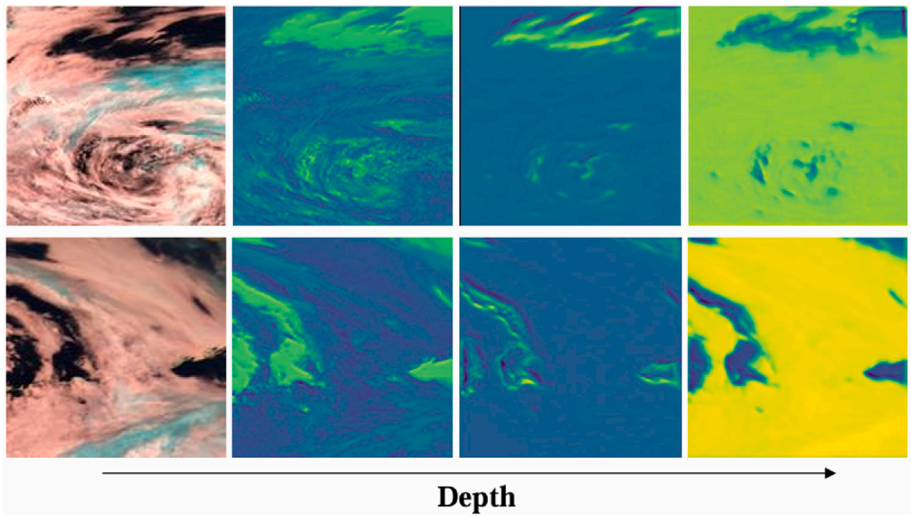

Fig. 5. Feature maps at different depths in the SR network. The increase in depth of the SR network leads to the loss of water cloud features. This will increase the difficulty of HR image reconstruction.

3.4. Multi-scale feature aggregation block

Attention mechanisms can force the network to focus on critical information and ignore irrelevant information. Channel attention and self-attention are applied in DASA to obtain more information features. However, these methods are still inadequate for the absorption of long-range dependence and local information, they usually consider attentional mappings with fixed receptive fields. In addition, the operation mode of MLP is not suitable for spatial information-rich data like images. MLP only adopts simple linear mapping for feature interaction, and for satellite cloud images, it is difficult for the network to accurately process the feature information of the areas with cloud occlusion and terrain edge. Therefore, MFAB is designed to compensate for the lack of attention mechanism. As shown in Fig. 6, MFAB consists of shift convolution and multi-scale interactive operation.

Shift Convolution. The shift convolution can benefit local pixels

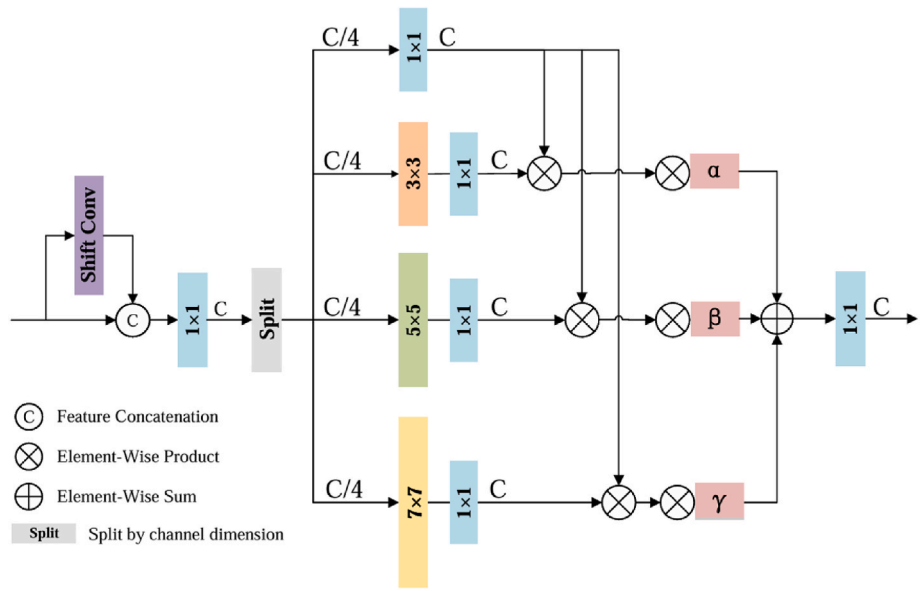

Fig. 6. Illustration of MFAB, which utilizes richer regional features, allowing them to interact spatially with the target region and converge adaptively to mitigate the drawbacks of satellite cloud maps, such as LW, blurring, and cloud occlusion.

participating in shifting features among different channel groups. TCSR (Gang Wu et al., 2023) and EMT (Ling Zheng et al., 2023) have proved the validity of shift convolution. As shown in Fig. 7, the input features are first split uniformly into N thinner groups along the channel dimension. Then each grouped feature is shifted in different directions with the shift step s. Authors set the constant value 0 as the default padding of the borders.

Multi-Scale Interactive Operation. Multi-scale interactive operation is utilized to further establish the association between features in neighboring regions. Given the input feature $F_x \in R^{B\times C\times H\times W}$, it first passes a 1×1 convolution and then is split equally into four parts in the channel dimension, named F_1, F_2, F_3, $F_4 \in R^{B\times \frac{C}{4}\times H\times W}$. Then F_1 is recovered to the number of channels of C by a 1×1 convolution, and the F_2, F_3, F_4 are fed into group convolutions with different convolution kernels for multi-scale feature extraction. The process of MFAB can be expressed as:

$$\begin{aligned} F_x &= H_{cat}(H_{sc}(F_{in}), F_{in}) \\ \widetilde{F}_x &= c_1(F_x) \end{aligned} \tag{9}$$

where $H_{cat}(\cdot)$ denotes feature concatenation, $H_{sc}(\cdot)$ denotes shift convolution. $c_1(\cdot)$ denotes 1×1 convolution. The features are then split by channel dimension and processed by multiscale convolutions:

$$F_1, F_2, F_3, F_4 = split(\widetilde{F}_x) \tag{10}$$

$$\begin{aligned} \widetilde{F}_1 &= u_1(F_1) \\ \widetilde{F}_2 &= u_2(f_{3\times 3}(F_2)) \\ \widetilde{F}_3 &= u_3(f_{5\times 5}(F_3)) \\ \widetilde{F}_4 &= u_4(f_{7\times 7}(F_4)) \end{aligned} \tag{11}$$

where $u_i(\cdot)(i\in 1,2,3,4)$ denotes 1×1 convolution used to expand the number of channels and interactive information. $split(\cdot)$ denotes the operation in which the feature is split on the channel dimension. $f_{3\times 3}$, $f_{5\times 5}$ and $f_{7\times 7}$ denote convolutions with different sizes. $\widetilde{F}_i \in R^{B\times C\times H\times W}(i\in 1,2,3,4)$ denotes the multi-scale features extracted by convolutions with different sizes. To achieve spatial interaction of feature information, $\widetilde{F}_1$ is multiplied element by element with $\widetilde{F}_2$, $\widetilde{F}_3$, $\widetilde{F}_4$ respectively to obtain the features $Y_i \in R^{C\times H\times W}(i\in 1,2,3,4)$:

$$\begin{aligned} Y_1 &= \widetilde{F}_1 * \widetilde{F}_2 \\ Y_2 &= \widetilde{F}_1 * \widetilde{F}_3 \\ Y_3 &= \widetilde{F}_1 * \widetilde{F}_4 \end{aligned} \tag{12}$$

this type of spatial interaction avoids the calculation of quadratic complexity. Then the features are given weights and aggregated, the process is represented as:

$$Y = \alpha Y_1 + \beta Y_2 + \gamma Y_3 \tag{13}$$

Y denotes the features after adaptive aggregation. α, β, γ are learnable weights, and $\alpha+\beta+\gamma = 1$. Finally, The dimension of Y is reduced to the initial number of channels by 1×1 convolution:

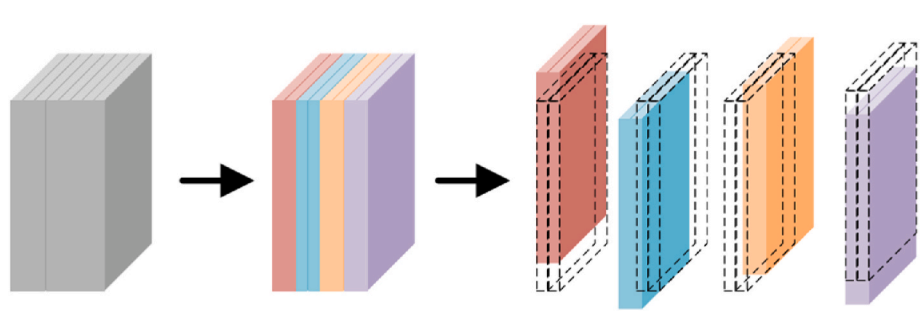

Fig. 7. Illustration of the shift convolution, which can benefit local pixels to participate in shifting features among different channel groups.

$$F_y = c_2(Y) \tag{14}$$

$F_y \in R^{C\times H\times W}$ indicates the output of MFAB. $c_2(\cdot)$ denotes 1×1 convolution to aggregate features across channels. MFAB realizes spatial information interaction and aggregation of features with fewer parameters and computational complexity. This approach facilitates nearby cloud-obscured areas to accomplish SR tasks through a wider sensory field. In addition, MFAB is a plug-and-play module.

DASA effectively captures features at different depths through dual-path convolutions and uses channel attention enhancement to help explore the correlation between features in three dimensions: space, channel, and depth. One path retains feature information through 1D convolution, and the other path extracts deeper features through a convolution group with channel attention to avoid the loss of features such as water clouds. MFAB uses rich regional features to enable the model to interact spatially with the target area and converge adaptively, alleviating the adverse effects of satellite cloud image blur and cloud occlusion on SR tasks. The collaborative working mechanism of DASA and MFAB improves the SR effect, and the proposed model has fewer parameters and computational complexity. This approach effectively handles satellite cloud image-specific challenges, such as the translucency and irregular shapes of clouds and the effects of terrain and waves on SR tasks.

3.5. Training Objective

The final parameter optimization model is obtained by minimizing the loss between the final merged reconstructed HR image and the ground truth. The L_1 loss is applied to train the network:

$$L(\theta) = \frac{1}{N}\sum_{i=1}^{N}\left\|I_H^i - I_{GT}^i\right\|_1 \tag{15}$$

where θ represents the parameter set of the network. N is the total number of training samples. I_H is the reconstructed HR image of the network I_{GT} is the ground truth. Although minimizing L_2 loss can maximize the peak signal-to-noise ratio (PSNR), L_1 loss provides better robustness for network training (Bee Lim et al., 2017).

4. Experiments and analysis

4.1. Datasets and evaluation metrics

The training of SR networks requires a set of HR images and their corresponding LR images, few meteorological satellite cloud datasets are currently available specifically for SR research (Bai Cong et al., 2021). Zhang et al. produced the NCCI dataset (Jinglin Zhang et al., 2021) for SR studies. NCCI dataset based on the Himawari-8 satellite HR data (80° E−160° W and 60° S-60° N). The authors selected satellite data from the Northern Hemisphere for the summer of 2017–2018 to generate satellite cloud images and extracted albedo data for band-5 (1.6 μm), band-4 (0.86 μm), and band-3 (0.64 μm) to generate natural color RGB composite images. On the NCCIs, ice clouds and water clouds with different properties and textures can be effectively distinguished. High clouds composed of ice particles, snow, and sea ice have low albedo on band-5 and high albedo on band-4 and band-3, so they appear blue on the image. Low clouds composed of water droplets have high albedo on all three bands and appear white or grey. In addition, the ocean is black, the desert is brown, and the vegetation is green in the image due to their albedo properties.

The NCCI dataset is used to train networks. Each NCCI is divided into multiple patches of size 600 × 600. The patches are manually selected and classified to obtain the HR satellite cloud image dataset containing five categories (ice clouds, water clouds, ocean, desert, and vegetation). The HR images are down-sampled using bicubic interpolation to obtain the corresponding LR images. The NCCI dataset contains a total of 5000 pairs of images. The number of images in each class is the same. 3000 pairs of images are randomly selected for training, and 30 pairs are used for validation during training. Authors build five test sets of five categories, each containing 26 pairs of images from outside the training set, as shown in Table 2. Samples of the NCCI dataset are exhibited in Fig. 8.

In addition, authors also train DPAT with the DIV2K dataset to verify its generalization ability. DIV2K dataset, based on natural images contains a variety of things with HR, clear contours of things, and obvious details. In contrast, the NCCI dataset is based on satellite cloud images and contains geography as well as clouds and fog. Some terrains do not have accurate and clear contours, while irregularly shaped and inconsistent transparency clouds can cause varying degrees of impact. Meanwhile, the resolution of NCCI dataset is only a quarter of that of DIV2K dataset. So it is difficult to motivate the full performance of the models during the training on NCCI dataset, and the SR visual effect is not obvious. Due to the limited resolution of the NCCI dataset, the model is difficult to train with ×8 scaling on it. To compare performance with other typical SR methods, authors evaluate the experimental results by the peak signal-to-noise ratio (PSNR) and the structural similarity index metrics (SSIM) (Zhou Wang et al., 2004) on the Y channel only in the converted YCbCr space. In addition, "Params" is used to indicate the total number of learnable parameters, and "FLOPs" is used to denote the number of floating point operations. The "FLOPs" are measured in SR of images of size 3 × 20 × 20 × 4.

This study adopted a series of strategies to prevent overfitting and ensure the balance of the dataset. The authors manually screened and classified the NCCI dataset to ensure consistency in the number of images in five different categories. Through data augmentation techniques, the authors downsampled the HR images to generate corresponding sets

Table 2
Datasets description.

Datasets	Information	Number	Data Balancing Technology
NCCI	Each natural-color cloud image was divided into multiple patches with a size of 600 × 600 pixels. The patches were manually selected and classified to obtain an HR satellite cloud image data set containing five categories (ocean, desert, vegetation, ice clouds, and water clouds).	5000	The HR images are down-sampled using bicubic interpolation to obtain the corresponding LR images.
DIV2K	Contains 1000 LR images with different degradation types, and the training data is 800 LR images. The validation data is 100 high-definition HR images, which are used to generate LR images. The test data is 100 diverse images, which are used to generate LR images.	1000	Evaluate the experimental results by the PSNR and the SSIM on the Y channel only in the converted YCbCr space.
Set5	A dataset consisting of 5 HD images with various scenes and textures, covering common challenges of image SR algorithms, such as detail enhancement, edge recovery, and comfort enhancement.	5	For testing
Set14	A larger dataset consisting of 14 HD images with different scenes and content, including natural scenery, indoor scenes, animals, and people, which better reflects the diversity and complexity of the real world.	14	For testing

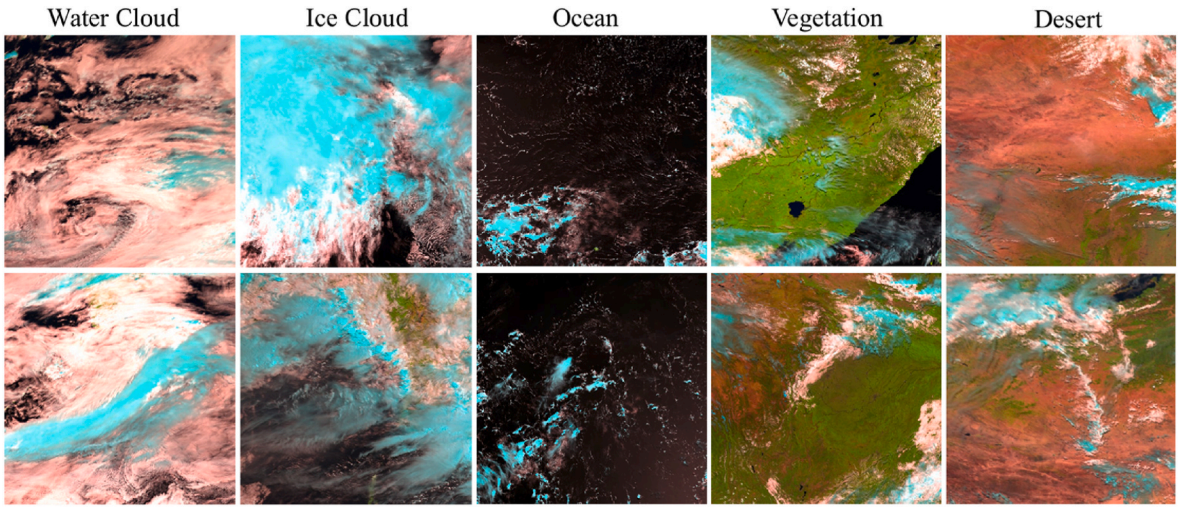

Fig. 8. The samples of NCCI dataset, which contains five categories (ice clouds, water clouds, ocean, desert, and vegetation).

of LR images. During the training process, each mini-batch randomly cropped color image patches of a specific size from the LR images as the input of the network, which increased the dataset's diversity and promoted the model's generalisation ability. The authors used convolution kernels of different sizes to aggregate multi-scale features, which significantly improved the robustness of the model to the cloud occlusion effect in satellite cloud images. The learning rate was halved after 200 epochs in each training cycle to optimise the model training further. This adaptive adjustment strategy helps the model to make more refined weight adjustments in the later stages of training and effectively avoid overfitting.

4.2. Implementation and training details

For DPAT, the number of MTBs is set to 6 and the number of MTLs in MTB is set to 6. The image reconstruction module has a sub-pixel convolution layer and a 1×1 convolution layer. The channel number of the intermediate feature is set to 64, and the GELU activation function is selected in the model. The length w_1 and width w_2 of the striped window in the DASA are set to 12 and 8, respectively.

During training, in each mini-batch, authors randomly crop 16 color patches with a specific size (96×96 for $\times 2$ and 192×192 for $\times\ 4$) from the LR images as input. The learning rate of all layers is initialized to 10^{-4} and decreased by half every 200 epochs, for a total of 800 epochs, the training-validation-test ratio is 8:1:1, as shown in Table 3. All experiments are performed using PyTorch 1.7.1 and Python 3.8 on NVIDIA GeForce GTX 3090 GPU.

4.3. Comparison with other methods

In order to reveal the effect of proposed network, several popular SISR networks are compared with DPAT, including RCAN (Yulun Z et al., 2004), DRLN (Saeed Anwar et al., 2019), SAN (D. Tao et al., 2019), HAN (Ben N et al., 2020), EDSR (Bee Lim et al., 2017), SwinIR (Jingyun L et al., 2021), and HAT (Xiangyu Chen et al., 2023). These networks are deep learning networks that have achieved state-of-the-art. The comparison results are presented in Table 4. Higher PSNR and SSIM values indicate better performance. Note that the number of parameters and FLOPs are adopted to measure the size of networks. It can be seen that the performance of the transform-based networks is higher than the CNN-based networks. On NCCI dataset, SwinIR achieves optimal metrics for both the 'Average' and the five satellite image categories, with DPAT being slightly below SwinIR and almost equal to HAT. In addition, Table 6 illustrates that DPAT has a much lower number of parameters and FLOPs than SwinIR. Although the number of parameters of HAT is lower than SwinIR, FLOPs are higher. The number of parameters in DPAT accounts for only 60% of the RCAN and 77% of the SwinIR. And DPAT has half the FLOPs of HAT. This demonstrates that the effects of lightweight design in network are obvious. Fig. 9 exhibits the visual comparison of these networks for $\times$ 4 SR on NCCI dataset.

In addition, DPAT is also trained with DIV2K dataset and tested on Set5 (Marco Bevilacqua et al., 2012), Set14 (Roman Zeyde et al., 2010), and BSD100 (David Martin et al., 2001) dataset to verify its generalization ability. The comparison results are presented in Table 7. DPAT achieves a balance between high accuracy and the number of parameters. Fig. 12 shows the visual comparison of the networks on Set5 and BSD100 for $\times$ 4 SR.

As shown in Fig. 9, a visual comparison of the processing results of DPAT and other SR methods on the NCCI dataset shows that SwinIR has significant advantages in enhancing image details, especially in depicting cloud edges and surface features, and can provide richer visual information. However, the method also shows slight signs of artificial processing in the processing of some areas, especially in the phenomenon of over-sharpening. DPAT performs more naturally in preserving the natural texture of clouds without introducing unnatural enhancement or blurring effects and excels in controlling ringing effects. In addition, DPAT also performs significantly better than other methods in colour reproduction and contrast balance, providing more realistic and visually more comfortable image output. Therefore, based on these advantages of DPAT in texture preservation and artefact control, this method is more suitable for SR tasks of satellite cloud images, especially in application scenarios that require in-depth analysis of complex meteorological phenomena. As shown in Fig. 12, in the test results on Set5 and BSD100 datasets, DPAT shows excellent performance in detail clarity, texture preservation, and color contrast. Although SwinIR provides some improvement in some details, it still needs further optimization to improve the naturalness of edge processing and color harmony. The qualitative analysis results confirm the potential of DPAT as an effective SR technique for processing satellite cloud images, especially in

Table 3
Experimental parameter settings.

Parameters	Values
Epochs	800
Training Batch size	8
Testing Batch size	1
Initial learning rate	0.0001
Optimizer weight decay	0.5
Image size($\times$ 2/$\times$ 4)	96×96/192×192
Training-validation-test ratio	8:1:1

Table 4
Effectiveness comparison (average PSNR/SSIM) of networks on different categories of NCCI datasets. 'Average' denotes the validation set, which contains 30 pairs of images in five categories. The other five test datasets each consisted of 26 pairs of images. The best and second best results are in red and blue.

Method	Scale	Average	Desert	Water	Ocean	Ice	Vegetation
Bicubic	×2	32.319/0.8971	34.538/0.9011	33.070/0.9004	28.526/0.8739	30.685/0.9135	32.682/0.8840
RCAN	×2	35.835/0.9397	37.605/0.9472	36.188/0.9360	31.675/0.9368	32.869/0.9357	36.361/0.9376
HAN	×2	35.701/0.9224	37.452/0.9187	35.297/0.9284	31.583/0.9279	32.765/0.9293	36.115/0.9281
EDSR	×2	35.511/0.9103	37.218/0.9047	35.014/0.9126	31.418/0.9152	32.601/0.9206	35.984/0.9048
DRLN	×2	35.804/0.9316	37.563/0.9394	36.147/0.9347	31.626/0.9305	32.843/0.9324	36.243/0.9362
SAN	×2	35.938/0.9458	38.691/0.9539	36.276/0.9439	31.745/0.9401	33.157/0.9447	36.338/0.9482
SwinIR	×2	36.101/0.9531	38.852/0.9608	36.521/0.9552	31.957/0.9408	33.483/0.9481	36.695/0.9538
HAT	×2	36.079/0.9487	38.843/0.9599	36.514/0.9546	31.887/0.9408	33.472/0.9470	36.687/0.9525
DPAT	×2	36.082/0.9495	38.861/0.9610	36.523/0.9554	31.889/0.9403	33.477/0.9479	36.700/0.9539
Bicubic	×4	28.253/07490	31.420/0.8002	29.259/0.7763	24.720/0.7225	26.479/0.7453	29.596/0.7556
RCAN	×4	30.481/0.8327	32.857/0.8492	30.962/0.8343	27.064/0.8218	28.188/0.8301	30.830/0.8255
HAN	×4	30.477/0.8311	32.829/0.8371	30.943/0.8323	27.053/0.8195	28.177/0.8316	30.837/0.8258
EDSR	×4	30.452/0.8154	32.817/0.8402	30.918/0.8299	27.002/0.8211	27.997/0.8305	30.826/0.8240
DRLN	×4	30.478/0.8326	32.867/0.8480	30.946/0.8330	27.066/0.8209	28.181/0.8355	30.849/0.8261
SAN	×4	30.484/0.8379	32.879/0.8509	30.958/0.8337	27.068/0.8219	28.198/0.8367	30.857/0.8264
SwinIR	×4	30.505/0.8427	32.916/0.8585	30.974/0.8362	27.076/0.8240	28.243/0.8408	30.863/0.8277
HAT	×4	30.501/0.8418	32.902/0.8557	30.964/0.8348	27.071/0.8237	28.238/0.8401	30.857/0.8271
DPAT	×4	30.510/0.8428	32.918/0.8586	30.970/0.8358	27.075/0.8238	28.247/0.8408	30.860/0.8272

Table 5
Weather satellite grayscale image experiment PSNR/SSIM results and Single epoch time (min).

Method	Scale	PSNR/SSIM	Single epoch time (min)
Bicubic	× 2	34.989/0.9081	8.630
RCAN	× 2	35.114/0.9102	8.477
HAN	× 2	35.531/0.9185	18.334
EDSR	× 2	35.415/0.9151	16.821
DRLN	× 2	35.223/0.9146	9.657
SAN	× 2	35.636/0.9233	4.968
SwinIR	× 2	35.806/0.9267	4.875
HAT	× 2	35.764/0.9265	5.054
DPAT	× 2	35.968/0.9312	1.965
Bicubic	× 4	38.983/0.9478	8.524
RCAN	× 4	39.066/0.9481	9.983
HAN	× 4	39.082/0.9488	20.534
EDSR	× 4	39.091/0.9490	18.184
DRLN	× 4	39.057/0.9485	10.319
SAN	× 4	39.096/0.9489	5.652
SwinIR	× 4	39.138/0.9492	4.258
HAT	× 4	39.115/0.9491	6.625
DPAT	× 4	39.135/0.9492	2.251

meteorological applications that require extremely high visual quality and image analysis accuracy.

Authors also examine three models with LAM. As shown in Fig. 13, it can be seen that RCAN tends to focus more on local information, while SwinIR is able to establish more remote dependencies and focus on more pixels, but does not make as much use of local information as RCAN. Both remote and local dependencies are important when recovering irregular features in satellite cloud images. Especially in areas disturbed by clouds, more information is required to assist in reconstruction. DPAT enhances the exploitation of local information while establishing remote dependencies.

To further demonstrate the adaptability of the DPAT mode in image super-resolution, we collected a dataset of 2500 satellite images that meet meteorological standards in real scenarios. As shown in Fig. 10, the data contains grayscale images of different weather conditions and geographic features at different resolutions.

We randomly selected 2200 images as training sets and 300 images as test sets and conducted experimental verification in Scale × 2 and × 4. We also explored the experimental efficiency and showed the time taken for a single epoch training, which intuitively represents the computational efficiency of the model. As shown in Table 5, in the case of Scale × 2, DPAT achieved the best results in both PSNR and SSIM, which are 35.968 and 0.9312, respectively. As shown in Fig. 11, its single epoch training time is also the shortest among all methods, at only 1.965 min. This shows that DPAT has high computational efficiency while maintaining high image quality.

In the case of Scale × 4, DPAT's PSNR and SSIM are 39.135 and 0.9492, respectively. Although it is not the best among all methods (SwinIR is slightly higher in PSNR), its single epoch training time is still the shortest, which is only 2.251 min. These results show that the DPAT model is adaptable to image SR tasks, especially in terms of computational efficiency. This is especially important for application scenarios that require rapid processing of large amounts of image data, such as meteorological satellite image analysis.

Fig. 14 illustrates the training and validation loss curves for DPAT alongside those of other state-of-the-art models. It is evident from the graph that DPAT exhibits a steady decrease in loss, indicating a robust learning process. Moreover, the validation loss for DPAT stabilizes at a lower value compared to other models, suggesting better generalization on unseen data.

Table 6
Quantitative comparison of the networks on NCCI dataset for × 2 SR and × 4 SR. The best and second best results are in red and blue.

Method	Scale	Params	FLOPs	PSNR
RCAN	×2	15.4M	49.0G	35.835
HAN	×2	15.9M	50.6G	35.701
EDSR	×2	40.7M	130.3G	35.311
DRLN	×2	34.4M	109.2G	35.783
SAN	×2	15.3M	48.9G	35.883
SwinIR	×2	11.5M	53.0G	36.101
HAT	×2	9.6M	54.7G	36.079
MetaSR	×2	14.5M	47.6G	36.902
DPAT	×2	8.8M	26.9G	36.082
RCAN	×4	15.6M	51.0G	30.481
HAN	×4	16.1M	52.5G	30.477
EDSR	×4	43.1M	160.3G	30.452
DRLN	×4	34.6M	111.2G	30.478
SAN	×4	15.9M	50.8G	30.484
SwinIR	×4	11.6M	53.2G	30.505
HAT	×4	9.6M	54.9G	30.501
MetaSR	×4	14.7M	47.7G	36.491
DPAT	×4	8.9M	27.2G	30.510

4.4. Ablation studies

In this section, several ablation studies are designed to validate DPAT design, long-term dependency modeling, and implementation details.

Study of window sizes. To enable the network to better establish long-term dependencies, self-attention is implemented on the striped window. Authors investigate the effect of window size on model performance by evaluating the PSNR of DPAT when using windows with different sizes on NCCI dataset for ×4 SR. The comparison results are presented in Table 8, and it can be seen that the increase in window size can improve the effect of SR, and the striped window will bring more performance improvement. When using the window size of 12× 12 or 24 × 6, the model takes too many resources and inference time, and the increase in accuracy is minimal. Therefore, the window size of 12× 8 is utilized for subsequent experiments.

Study of MTL. MTL can establish long-range dependencies and aggregate multi-scale features more efficiently. Unlike the original Transformer layer, MTL consists of DASA and MFAB. To verify the effectiveness of MTL, several variant models are designed, and SwinIR is used as the base model of the comparison. The SR results are reported in Table 9. DASA drastically reduces the number of parameters and FLOPs at the cost of less accuracy loss. Compared to the original Transformer layer, which directly processes all features with linear mapping, using convolution processing first and then linearly transforming features will reduce the number of parameters. The MFAB, which replaces the MLP, exhibits approximate performance with fewer parameters and computation cost. MFAB with shift convolution can aggregate richer multi-scale features without additional parameters and computational effort, and the PSNR grows by 0.2 dB and 0.002 dB on Set5 and NCCI respectively.

In addition, to further evaluate the model performance, the depth of the network is increased or decreased for comparison. As presented in Table 10, it can be seen that SwinIR requires more parameters and FLOPs to achieve the same performance as DPAT on NCCI dataset.

Study of DASA. The dual-path convolution is utilized to extract more feature information for self-attention. Weights α and β are assigned to the *queries* computed by each of the two paths, and $\alpha + \beta = 1$. Authors experimentally explore the most appropriate weight values. As Table 11 presents, the network performs best when α and β are 0.2 and 0.8 respectively, which increases by 0.01 dB compared to a single-path structure without channel attention and convolution.

In addition, inspired by HAT, channel attention is added to the network to promote self-attention to activate more pixels. Authors also verify the effect of different attention mechanisms. Table 12 demonstrates the results of three attention mechanisms combined with self-attention, including channel attention (Marco Bevilacqua et al., 2012), spatial attention (Roman Zeyde et al., 2010), and channel-spatial attention (David Martin et al., 2001). It can be seen that the channel attention promotes self-attention to activate more pixels to improve performance. Spatial attention instead inhibits the effect of

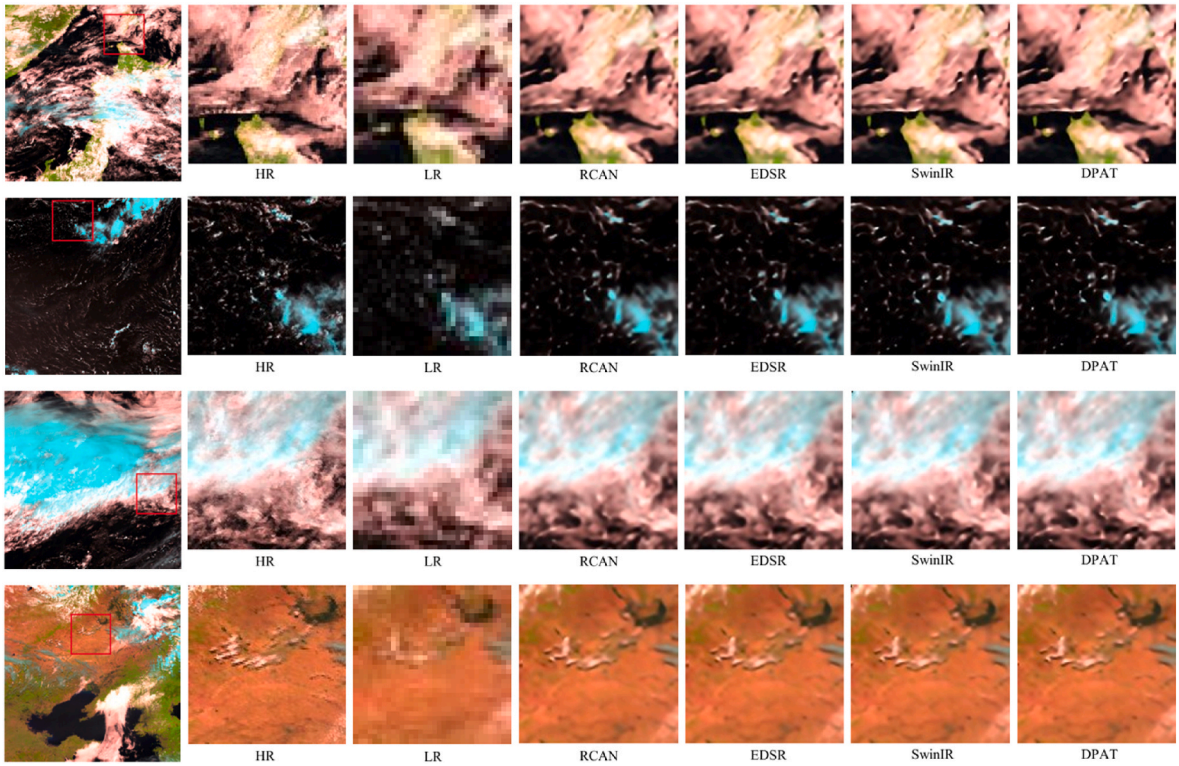

Fig. 9. Visual comparison of DPAT with other methods on NCCI dataset for ×4 SR. The visual effect of DPAT is better than the CNN-based networks and approximates SwinIR.

Table 7
Comparison of networks trained on DIV2K dataset and tested for × 4 SR. The best and second best results are in red and blue.

Method	Params	Set5	Set14	BSD100
RCAN	15.6M	32.63	28.87	27.77
HAN	16.1M	32.64	28.92	27.80
EDSR	43.1M	32.46	28.80	27.71
DRLN	34.6M	32.63	28.94	27.83
SAN	15.9M	32.64	28.92	27.78
SwinIR	11.6M	32.72	28.94	27.83
HAT	9.6M	32.71	28.92	27.80
MetaSR	14.7M	32.63	28.90	27.79
DPAT	8.9M	32.72	28.93	27.81

self-attention, and authors consider that these two types of attention conflict with each other when computed in the spatial dimension of the feature image.

Study of MFAB. MFAB replaces MLP to associate feature information under different sensory fields, which can mitigate the effect of occlusion by cloud in satellite cloud images. In MFAB, different scales correspond to different sizes of convolution kernel. Authors try three combinations of convolutional kernel sizes for 1-3-5, 3-5-7, 5-7-9, and 3-7-11. Table 13 presents the comparison results. It can be seen that larger convolution kernels cannot improve performance, and the combination of 3-5-7 sizes of convolution kernel is the most suitable.

4.5. Cross-validation

To further verify the robustness and generalization ability of the DPAT model, the authors implemented K-Fold cross-validation on the NCCI dataset. 5-Fold cross-validation was selected, that is, K = 5. The 5000 pairs of images in the NCCI dataset were evenly divided into five subsets. Each subset contains 1000 pairs of images, and the number of images in each category is consistent. In each iteration, 4 subsets (4000 pairs of images) are used for training, and the remaining 1 subset (1000 pairs of images) is used for verification.

Table 14 shows the average PSNR and SSIM values of 5 iterations and the standard deviation. The results show that the DPAT model exhibits small performance fluctuations on different data partitions, which

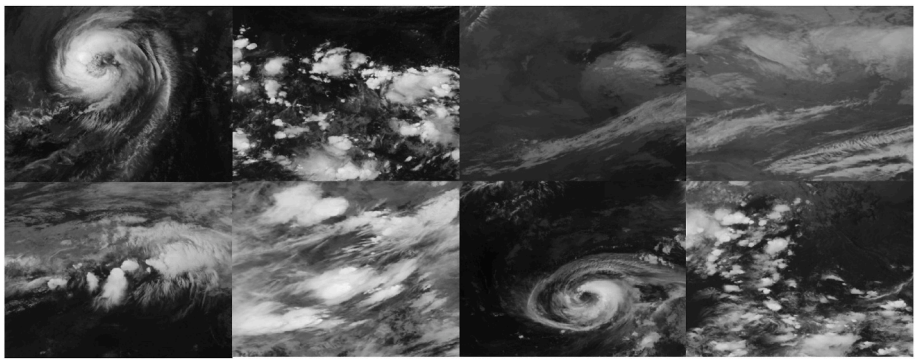

Fig. 10. The grayscale data of meteorological images in real scenes were collected, including different weather conditions and geographical information, to verify the adaptability of DPAT.

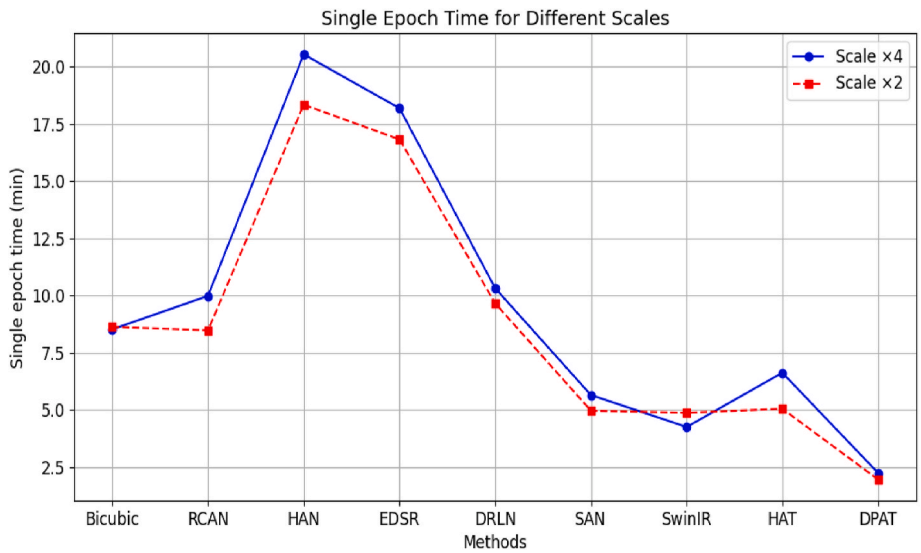

Fig. 11. Single epoch time for different scales.

verifies the stability and reliability of the model.

4.6. Discussion

The extensive experiments and analyses conducted in this study provide valuable insights into the performance and characteristics of the DPAT model in the context of meteorological satellite cloud image SR.

The introduction of the NCCI dataset is a significant contribution to the SR research community, providing a unique and challenging dataset tailored to the specific characteristics of satellite imagery. The composition of this dataset and its unique classes and inherent properties of albedo reflective surfaces provide a comprehensive platform for training and evaluating SR models such as DPAT. Comparative studies with other state-of-the-art methods, including RCAN, DRLN, SAN, HAN, EDSR, SwinIR, and HAT, show that DPAT is highly competitive. It achieves results comparable to the best-performing models, such as SwinIR, while significantly reducing the number of parameters and computational complexity of the model.

Ablation studies provide a deeper understanding of the inner workings of the model. The impact of window size, the role of DASA and MFAB, and the impact of different attention mechanisms are carefully examined. The results highlight the importance of balancing model complexity and performance gains and emphasize that not all increases in model complexity necessarily lead to proportional improvements in results. The K-Fold cross-validation results further confirm the robustness of the DPAT model. The consistency between the different folds indicates that the model does not overfit to a specific subset of the data and generalizes well across the entire dataset.

The visual comparison presented in Fig. 12 and the LAM analysis in Fig. 13 provides a qualitative perspective on the model performance. DPAT excels in recovering fine details and maintaining the structural integrity of the image, even when faced with the challenges of complex textures and patterns in satellite imagery. DPAT achieves high accuracy with a lightweight design, making it a promising candidate for practical applications with limited computational resources. The model can generalize to other datasets, as evidenced by its performance on the DIV2K dataset and standard benchmarks such as Set5, Set14, and BSD100, further highlighting its versatility. The DPAT model represents a significant advancement in satellite image SR, offering an excellent combination of efficiency, effectiveness, and generalization.

The DPAT presented in this study demonstrates significant parameter efficiency and computational complexity advantages compared to existing models such as SwinIR. The design choices made in DPAT aim not only to reduce the model's footprint but also to enhance its adaptability and generalizability across different imaging scenarios and hardware platforms. The lightweight strategy employed in DPAT is transferable and scalable due to its modular architecture, which allows for easy integration with other components and models. DASA mechanism and MFAB are designed to be independent modules that can be incorporated into various network configurations without significant modifications. This modularity ensures that the strategies used to reduce parameter count and computational complexity are not tied to a specific

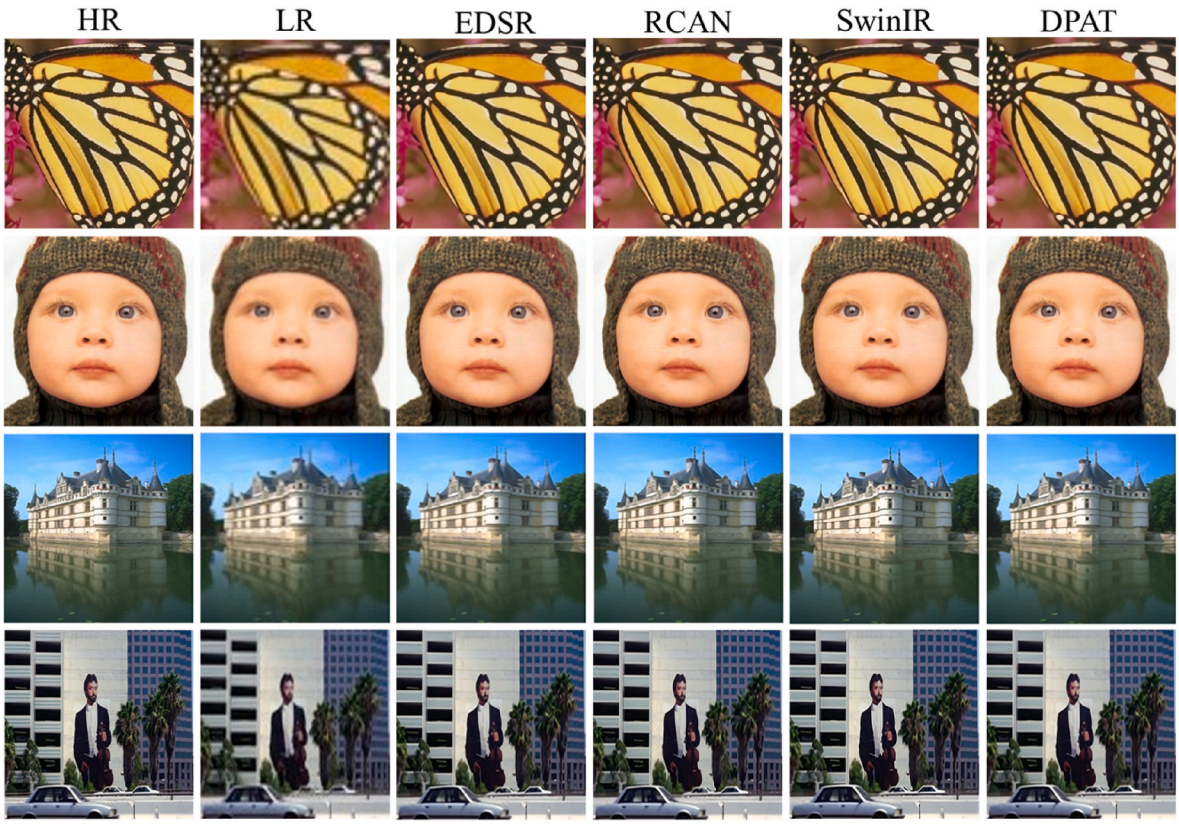

Fig. 12. Visual comparison of networks for × 4 SR on Set5 and BSD100. It can be seen that DPAT has excellent generalization ability on other datasets.

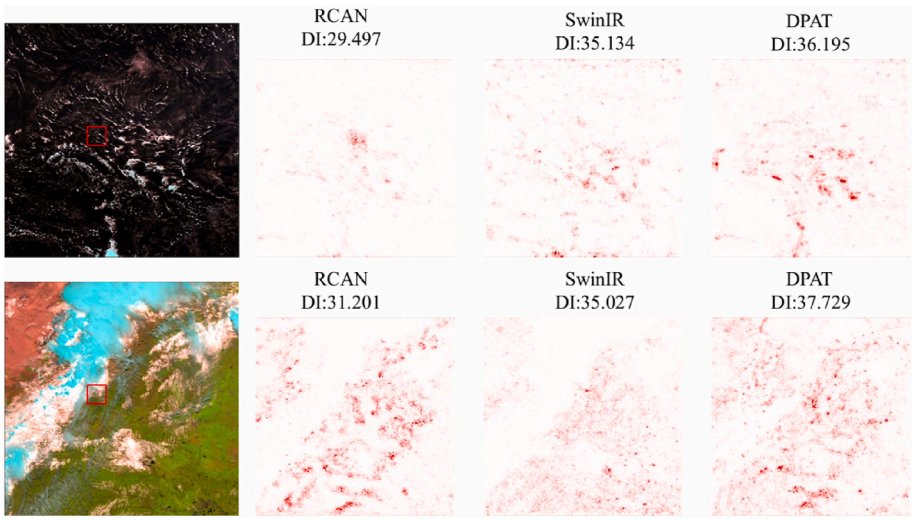

Fig. 13. Results of LAM attribution about different networks at × 4 image SR on the NCCI dataset. When reconstructing the patches marked with red boxes, the higher diffusion index (DI) indicates the greater range of contextual information involved, and the darker color indicates the greater degree of contribution.

model structure but can be applied broadly across different deep learning architectures. By focusing on the most informative features and reducing redundancy, DPAT optimizes the information flow through the network, which inherently reduces the need for extensive parameterization. Shift convolution in MFAB allows for efficient capture of spatial hierarchies with a minimal increase in parameters, thus maintaining a low computational overhead. The model's performance will be benchmarked against other state-of-the-art methods under identical computational constraints to evaluate its real-world advantages comprehensively.

In conclusion, based on a deep understanding of satellite cloud image SR's underlying computational and informational requirements. The transferability and scalability of these strategies, along with their potential to reduce inference time, position DPAT as a promising candidate for practical deployment in various satellite imaging applications.

5. Conclusion

In this work, authors explore the effect of SR on satellite cloud images. The complex geography and irregularly shaped clouds in satellite

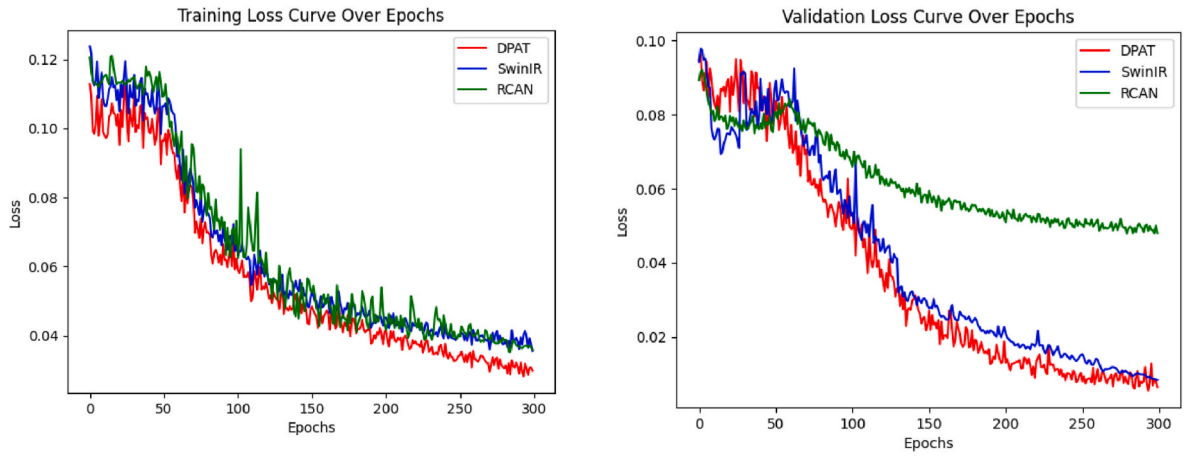

Fig. 14. The training and validation loss curves for DPAT alongside those of other state-of-the-art models, including SwinIR and RCAN.

Table 8
Quantitative comparison of different window sizes on NCCI dataset for × 4 SR.

Size	(8,8)	(16,4)	(12,8)	(12,12)	(24,6)
PSNR	30.493	30.501	30.510	30.508	30.507

Table 9
Ablation studies on DASA and MFAB in MTL for × 4 SR. SwinIR is utilized as the base model. SC denotes shift convolution.

Params	FLOPs	DASA	MFAB (w/o SC)	MFAB (w SC)	NCCI	Set5
11.6M	53.2G				30.505	32.72
9.5M	33.7G	✓			30.496	32.69
10.7M	46.8G		✓		30.501	32.70
10.7M	46.8G			✓	30.505	32.70
8.9M	27.2G	✓	✓		30.508	32.72
8.9M	27.2G	✓		✓	**30.510**	**32.73**

Table 10
Comparisons of networks for × 4 SR. † indicates that the model has reduced or increased the number of layers.

Method	Params	FLOPs	NCCI	Set5
SwinIR	11.6M	53.2G	30.505	32.72
†SwinIR	9.1M	42.2G	30.496	32.69
DPAT	8.9M	27.2G	30.510	32.73
†DPAT	11.6M	32.8G	30.515	32.74

Table 11
Comparison of different weights on paths with convolution groups on NCCI dataset for × 4 SR.

Weight(α)	0	0.2	0.5	0.8	1.0
PSNR	30.489	30.494	30.501	30.510	30.506

Table 12
Comparison of different attention mechanisms in DASA on NCCI dataset for × 4 SR.

Attention mechanisms	Params	PSNR
No attention	8.8M	30.492
Channel attention	8.9M	30.510
Spatial attention	9.0M	30.496
Channel-spatial attention	8.9M	30.497

cloud images make it difficult for SR models to perform their full performance. DPAT is designed to cope with the difficulties in the SR task of satellite cloud images. In DPAT, DASA is designed to utilize dual-path convolution and channel attention to benefit the self-attention mechanism to establish the dependency of features in three dimensions: space, depth, and channel. In addition, to establish the long-range dependence of features in images more efficiently, the striped window is introduced for self-attention mechanism. To better process the spatial information of the image and reduce the effect of cloud occlusion, the MLP in the original Transformer layer is replaced by MFAB. MFAB allows the model to aggregate sufficient information from more receptive fields to accomplish the SR task. To verify the validity of the proposed network, authors train DPAT and other popular networks on NCCI dataset. The evaluation results suggest that DPAT achieves high performance with the minimum number of parameters and computing time. DPAT takes only 77% of the parameters of SwinIR to achieve an approximate effect with SwinIR. In addition, DPAT also achieves a balance of high performance and low parameters on DIV2K dataset. In the future, authors will explore multi-scale feature fusion technology, fully use cloud information of different resolutions, continue to study model compression technologies such as model pruning and quantization, and reduce the model size and computational requirements of DPAT. In addition, the authors will also explore the application potential of DPAT technology in complex scenarios such as human action recognition, intention perception, and industrial recognition, in order to significantly improve the accuracy of image super-resolution and image perception tasks.

Table 13
Comparison between different combinations of kernel sizes in MFAB on NCCI dataset for × 4 SR.

Kernel Sizes	Params	FLOPs	PSNR
1-3-5	8.8M	26.1G	30.496
3-5-7	8.9M	27.2G	30.510
5-7-9	9.1M	28.6G	30.507
3-7-11	9.0M	29.1G	30.510

Table 14
K-Fold cross-validation results on NCCI dataset for 4x SR.

Fold	Average PSNR	Average SSIM	PSNR Std Dev	SSIM Std Dev
1	30.508	0.9495	0.05	0.0005
2	30.496	0.9490	0.06	0.0006
3	30.515	0.9501	0.04	0.0004
4	30.497	0.9489	0.05	0.0005
5	30.510	0.9500	0.04	0.0004
Mean	30.505	0.9495	0.05	0.0005

CRediT authorship contribution statement

Qinghui Chen: Conceptualization, Data curation, Formal analysis, Writing – review & editing, Methodology, Resources. **Lunqian Wang:** Writing – review & editing, Writing – original draft, Methodology, Formal analysis, Data curation, Conceptualization. **Zekai Zhang:** Supervision, Formal analysis. **Xinghua Wang:** Investigation, Formal analysis. **Weilin Liu:** Software, Resources, Data curation. **Bo Xia:** Data curation, Conceptualization. **Hao Ding:** Investigation, Formal analysis. **Jinglin Zhang:** Writing – review & editing, Visualization, Project administration, Data curation. **Sen Xu:** Validation, Supervision. **Xin Wang:** Supervision.

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.

Acknowledgments

This work was supported in part by the National Key Research and Development Program of China under Grant 2022YFB4500602, in part by the Distinguished Young Scholar of Shandong Province under Grant ZR2023J0025, in part by the Taishan Scholars Program under Grant tsqn 202211290, in part by the Key Research and Development Program of Shandong Province of China under Grant 2023CXGC010112, and in part by the Major Basic Research Projectsof Shandong Province under Grant ZR2022ZD32.

Data availability

The NCCI data set can be downloaded from online (http://doi.org/10.5281/zenodo.3902586). The DIV2K data set can be downloaded from online https://data.vision.ee.ethz.ch/cvl/DIV2K/).

References

Agnihotri, J., Behrangi, A., Tavakoly, A., Geheran, M., Farmani, M.A., Niu, G.Y., 2023. " Higher frozen soil permeability represented in a hydrological model improves spring streamflow prediction from river basin to continental scales,". Water Resour. Res. 59 (4), e2022WR033075.
Ahire, R.B., Patil, V., 2013. Overview of satellite image resolution enhancement techniques. Wireless and Optical Communications Networks 1–4.
Akkem, Y., Biswas, S.K., Varanasi, A., 2023. Smart farming using artificial intelligence: a review. Eng. Appl. Artif. Intell. 120, 105899.
Alexey, D., Lucas, B., Alexander, K., Dirk, W., Xiaohua, Z., Thomas, U., Mostafa, D., Matthias, M., Georg, H., Sylvain, G., Jakob, U., Neil, H., 2021. An image is worth 16x16 words: transformers for image recognition at scale. International Conference on Learning Representations.
Anwar, Saeed, Barnes, Nick, 2019. Densely residual laplacian super resolution. IEEE Trans. Pattern Anal. Mach. Intell. 44, 1192–1204.
Bai, Cong, Zhang, Minjing, Zhang, Jinglin, Zheng, Jianwei, Chen, Shengyong, 2021. LSCIDMR: large-scale satellite cloud image database for meteorological research. IEEE Transactions on Cybernetics. 52 (11), 12538–12550.
Ben, N., Weilei, W., Wenqi, R., Xiangde, Z., Lianping, Y., Shuzhen, W., Kaihao, Z., Xiaochun, C., Haifeng, S., 2020. Single image super-resolution via a holistic attention network. European Conference on Computer Vision 11211, 191–207.
Bevilacqua, Marco, Roumy, Aline, Guillemot, Christine, Line Alberi-Morel, Marie, 2012. Low-complexity single-image super-resolution based on nonnegative neighbor embedding. In: British Machine Vision Conference, pp. 1–10.
Chan, Kelvin CK., Wang, Xintao, Yu, Ke, Dong, Chao, Loy, Chen Change, 2021. BasicVSR: the search for essential components in video super-resolution and beyond. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 4947–4956.
Chao, D., Chen, C.L., Kaiming, H., Xiaoou, T., 2016. Image super-resolution using deep convolutional networks. IEEE Trans. Pattern Anal. Mach. Intell. 99, 1–307.
Chen, X., Wang, X., Zhou, J., Qiao, Y., Dong, C., 2023. Activating more pixels in image super-resolution transformer. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. 22367–22377.
Da Wang, Y., Armstrong, R., Mostaghimi, P., 2019. Super resolution convolutional neural network models for enhancing resolution of rockmicroct images. Computing Research Repository arXiv preprint arXiv:1904.07470.
Demirel, H., Anbarjafari, G., 2011. Discrete wavelet transform-based satellite image resolution enhancement. IEEE Trans. Geosci. Rem. Sens. 49 (6), 1997–2004.
Farmani, M.A., Behrangi, A., Gupta, A., Tavakoly, A., Geheran, M., Niu, G.Y., 2024. What Are the Key Soil Hydrological Processes to Control Soil Moisture Memory? EGUsphere, pp. 1–28, 2024.
Garg, A., Naidu, S.V., Ahmed, T., Yahia, H., Singh, D., 2014. Wavelet based resolution enhancement for low resolution satellite images. International Conference on Industrial and Information Systems, pp. 1–5.
Gu, Jinjin, Dong, Chao, 2021. Interpreting Super-resolution Networks with Local Attribution Maps. Computer Vision and Pattern Recognition, pp. 9199–9208.
Henderson, B.G., Chylek, P., 2005. The effect of spatial resolution on satellite aerosol optical depth retrieval. EEE transactions on geoscience and remote sensing 43 (9), 1984–1990.
Jiang, J., Hu, R., Wang, Z., Han, Z., S, 2013. Dong Manifold regularized sparse support regression for single image super-resolution ICASSP. IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings 1429–1433.
Jin, W., Fu, R., Ye, M., 2014. Wavelet based resolution enhancement for low resolution satellite images. International Conference on Industrial and Information Systems, pp. 1–5.
Jingyun, L., Jiezhang, C., Guolei, S., Kai, Z., Luc, V.G., Radu, T., 2021. SwinIR: image restoration using Swin transformer. IEEE International Confer Ence on Computer Vision, pp. 1833–1844.
Kabir, S., Leigh, L., Helder, D., 2020. Vicarious methodologies to assess and improve the quality of the optical remote sensing images: a critical review. Rem. Sens. 12 (24), 4029.
Kawai, Y., Kawamura, H., Takahashi, S., Hosoda, K., Murakami, H., Kachi, M., Guan, L., 2006. Satellite-based high-resolution global optimum interpolation sea surface temperature data. J. Geophys. Res.: Oceans 111 (C6).
Keshk, Hatem Magdy, Yin, Xu-Cheng, 2017. Satellite super-resolution images depending on deep learning methods: a comparative study. In: International Conference on Signal Processing, pp. 1–7.
Kim, W.H., Kim, J.N., 2010. Cloud image resolution enhancement method using loss information estimation. Signal Processing and Multimedia 115–120.
Kun, Yuan, Guo, Shaopeng, Liu, Ziwei, Zhou, Aojun, Yu, Fengwei, Wu, Wei, 2021. Incorporating convolution designs into visual transformers. IEEE International Conference on Computer Vision, pp. 559–568.
Lepcha, D.C., Goyal, B., Dogra, A., S, H., 2022. Wang an efficient medical image super resolution based on piecewise linear regression strategy using domain transform filtering Concurr. Comput. Times 34 (20), e6644. https://doi.org/10.1002/CPE.6644.
Li, Feng, Bai, Huihui, Zhao, Yao, 2020. FilterNet: adaptive information filtering network for accurate and Fast image super-resolution. IEEE Trans. Circ. Syst. Video Technol. 30, 1511–1523.
Li, Y., Zhou, Y., Zhang, Y., et al., 2022. DKDFN: domain knowledge-guided deep collaborative fusion network for multimodal unitemporal remote sensing land cover classification. ISPRS J. Photogrammetry Remote Sens. 186, 170–189.
Li, Yawei, Fan, Yuchen, Xiang, Xiaoyu, Demandolx, Denis, Ran jan, Rakesh, Timofte, Radu, Van Gool, Luc, 2023. Efficient and explicit modelling of image hierarchies for image. abs/2303, 00748.
Lim, Bee, Son, Sanghyun, Kim, Heewon, Nah, Seungjun, Lee, Kyoung Mu, 2017. Enhanced deep residual networks for single image super resolution. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp. 136–144.
Martin, David, Fowlkes, Charless, Tal, Doron, Malik, Jitendra, 2001. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In: ICCV.
Mei, Yiqun, Fan, Yuchen, Zhou, Yuqian, 2021. Image Super-resolution with Non-local Sparse Attention, vol. 11211. Computer Vision and Pattern Recognition, pp. 3517–3526.
Muhammad, H., Greg, S., Norimichi, U., Shakhnarovich, G., 2018. Deep back projection networks for super-resolution. Computer Vision and Pattern Recognition 1664–1673 abs/1803.02735.
Safari, K., Imani, F., 2022. A Novel Fuzzy-BELBIC Structure for the Adaptive Control of Satellite attitude[C]//ASME International Mechanical Engineering Congress and Exposition, vol. 86656. American Society of Mechanical Engineers, V003T04A033.
Seydi, S.T., Boueshagh, M., Namjoo, F., Minouei, S.M., Nikraftar, Z., Amani, M., 2024. A hyperspectral change detection (HCD-Net) framework based on double stream convolutional neural networks and an attention module. Remote Sensing 16 (5), 827.
Shi, Jinpeng, Li, Hui, Liu, Tianle, Liu, Yulong, Zhang, Mingjian, Zhu, Jinchen, Zheng, Ling, Weng, Shizhuang, 2023. Image super-resolution using efficient striped window transformer. arXiv:, 2301, 09869.
Sun, Y., Deng, K., Ren, K., Liu, J., Deng, C., Jin, Y., 2024. Deep learning in statistical downscaling for deriving high spatial resolution gridded meteorological data: a systematic review. ISPRS J. Photogrammetry Remote Sens. 208, 14–38.
Tao, D., Jianrui, C., Yongbing, Z., Shu-Tao, X., Lei, Z., 2019. Second-Order Attention Network for Single Image Super-resolution, vol. 11211. Computer Vision and Pattern Recognition, pp. 11065–11074.
Tuna, C., Unal, G., Sertel, E., 2018. Single-frame super resolution of remote sensing images by convolutional neural networks. Int. J. Rem. Sens. 39 (8), 2463–2479.
Umehara, K., Ota, J., Ishida, T., 2018. Application of super-resolution convolutional neural network for enhancing image resolution in chestct. IEEE Trans. Geosci. Rem. Sens. 31 (4), 441–450.
Wang, Zhou, Conrad Bovik, Alan, Sheikh, Hamid Rahim, Simoncelli, Eero P., 2004. Image quality assessment: from error visibility to structural similarity. IEEE Trans. Image Process. 13 (4), 600–612.

Wang, Yifan, Wang, Lijun, Wang, Hongyu, Li, Peihua, 2018. RAN: resolution-aware network for image super-resolution. IEEE Transac Tions on Circuits and Systems for Video Technology, 1-1.
Wi, S., Steinschneider, S., 2022. Assessing the physical realism of deep learning hydrologic model projections under climate change. Water Resour. Res. 58 (9), e2022WR032123.
Wu, Haiping, Xiao, Bin, Codella, Noel, Liu, Mengchen, Dai, Xiyang, Lu, Yuan, Zhang, Lei, 2021. CvT: introducing convolutions to vision transformers. IEEE International Conference on Computer Vision, pp. 22–31.
Wu, Gang, Wu, Jiang, Junjun, Bai, Yuanchao, Liu, Xianming, 2023. Incorporating transformer designs into convolutions for lightweight image super resolution. arXiv preprint 2023 arXiv:2303.14324.
Xiao, Tete, Singh, Mannat, Mintun, Eric, Darrell, Trevor, Dollár, Piotr, Girshick, Ross, 2021. Early convolutions help transformers see better. Conference on Neural Information Processing Systems, pp. 30392–30400.
Xie, Chao, Zeng, Weili, Lu, Xiaobo, 2018. Ast single image super resolution via deep network with component learning. IEEE Trans 12 actions on Circuits and Systems for Video Technology 29, 3473–3486.
Xu, W., Guangluan, X., Wang, Y., Sun, X., Lin, D., Yirong, W., 2018. High quality remote sensing image super-resolution using deep memory connected network. IEEE International Geoscience and Remote Sensing Symposium 8889–8892.
Xu, M., Deng, F., Jia, S., Jia, X., Plaza, A.J., 2022. Attention mechanism-based generative adversarial networks for cloud removal in Landsat images. Remote sensing of environment 271, 112902.
Yang, J., Wright, J., Huang, T.S., Ma, Y., 2010. Image super-resolution via sparse representation. IEEE Trans. Image Process. 19 (11), 2861–2873.
Yang, X., Xie, T., Liu, L., Zhou, D., 2021. Image super-resolution reconstruction based on improved Dirac residual network Multidimensional. Systems and Signal Processing 32 (4), 1065–1082. https://doi.org/10.1007/S11045-021-00773-0, 2021 32:4Mar.
Ying, Z., Randi, F., Wen, Y., Feng, Z., Wei, J., 2016. A method of infrared nephogram super-resolution based on structural group sparse representation. Opto-Electronic Eng. 43 (12), 126–132.
Ying, Tai, Yang, Jian, Liu, Xiaoming, 2017. Image super-resolution via deep recursive residual network. Computer Vision and Pattern Recogni tion 2017, 2790–2798.
Yulun, Z., Kunpeng, L., Kai, L., Lichen, W., Bineng, Z., Yun, F., 2018. Image super resolution using very deep residual channel attention networks. European Conference on Computer Vision 11211, 294–310.
Ze, L., Yutong, L., Yue, C., Han, H., Yixuan, W., Zheng, Z., Stephen, L., Baining, G., 2021. Swin transformer- hierarchical vision transformer using shifted windows. IEEE International Conference on Computer Vision, pp. 9992–10002.
Zeyde, Roman, Elad, Michael, Protter, Matan, 2010. Low-On single image scale-up using sparse-representations. In: International Conference on Curves and Surfaces, pp. 711–730.
Zeyde, R., Elad, M., Protter, M., 2012. Nephogram super-resolution algorithm using over-complete dictionary via sparse representation. Yaogan Xuebao-. Journal of Remote Sensing 16 (2), 275–285.
Zhang, Jinglin, Yang, Zhipeng, Jia, Zhaoying, Bai, Cong, 2021. Super resolution imaging with A deep multipath network for the reconstruction of satellite cloud images. Earth Space Sci. 559–568.
Zhao, X., Zhang, Y., Zhang, T., Zou, X., 2019. Channel splitting network for single mr image super-resolution. IEEE Trans. Image Process. 28 (11), 5649–5662.
Zheng, Ling, Zhu, Jinchen, Shi, Jinpeng, Weng, Shizhuang, 2023. Efficient mixed transformer for single image super-resolution. CoRR abs/2305.11403.
Kunchang, L., Yali, W., Junhao, Z., Peng, G., Guanglu, S., Yu, L., Hongsheng, L., Yu, Q., 2023. UniFormer: unifying convolution and self-attention for visual recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence 45 (10), 12581–12600.