Multi-scale adaptive weighted network for polarization computational imaging super-resolution

Guoming Xu\textsuperscript{1,2,4} · Jie Wang\textsuperscript{1} · Lei Zhang\textsuperscript{3} · Jian Ma\textsuperscript{1,2} · Yong Wang\textsuperscript{4} · Jiaqing Liu\textsuperscript{1}

Received: 5 July 2022 / Accepted: 30 August 2022 / Published online: 14 October 2022
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2022

Abstract
Due to the pixel limit of the polarization imaging detector and the object detection conditions, the spatial resolution of the object polarization image in actual imaging detection applications is generally low. Convolutional neural networks (CNNs) have been introduced to image super-resolution (SR). However, these methods rarely explore the internal connections between local spatial components and multi-scale feature maps in different receptive fields. To this end, we propose a multi-scale adaptive weighted network (MSAWN) for polarization computational imaging super-resolution to gain superior reconstruction performance. Computational imaging methods centered on information acquisition and interpretation can obtain high-resolution images that are superior to imaging systems. Specifically, we use a limited amount of memory and computational power even with multi-scale and multi-level polarization information. Second, a spatial pyramid structure based on the space-channel attention mechanism is designed to effectively adjust the feature weight of polarization information. Third, we adopt an adaptive weight unit to reduce redundant network branches and parameters. Particularly, we design an innovative reconstruction layer with inputs coming from multiple paths by means of sub-pixel convolution. The experimental results show that the proposed method achieves better reconstruction accuracy and visual effect, and the objective evaluation indexes such as PSNR and SSIM are significantly improved.

1 Introduction

Polarization is one of the essential properties of light. The differences or changes in physical properties of aerosol or object can be represented by polarization characteristics in imaging detection. The high-dimensional polarization characteristics can effectively improve the contrast between object and background, which lays the foundation for inversion of object spatial structure and enhance the recognition of object in the cluttered background \cite{1}. Polarization imaging technology is an imaging technology that acquires the polarization state of light waves reflected on the surface of the object \cite{2}. These characteristics make polarization imaging detection widely used in military object recognition and tracking, medical image processing, machine vision, geographic information analysis, remote sensing image processing and other fields \cite{3, 4}.

The traditional time-sharing polarization imaging technology cannot obtain 4 Stokes parameters at the same time. Spatial modulation full polarization imaging technology is a new polarization imaging technology developed after the traditional time-sharing and simultaneous polarization imaging technology \cite{5–8}. In recent years, many scholars have...
carried out in-depth research on polarization image imaging technology. Durán et al. [9] used spatial light modulator and Stokes polarimeter to realize single-pixel polarization imaging, and obtain the object’s Stokes parameter image. Zhang et al. [7] improved the reconstruction algorithm through sum difference reconstruction, but there is still an urgent need for a fully polarized imaging system with high resolution in practical application. Due to the influence of imaging distance and atmospheric disturbance, the limit resolution of the image projected on the focal plane is greatly reduced (much smaller than the optical system diffraction limit resolution [10]), resulting in a lower spatial resolution of the polarized image. Moreover, the spatial resolution of the image is limited by the number of detector pixels. The high resolution (HR) image is of great significance and value for object detection accuracy. For this reason, without replacing the hardware imaging system, the image super-resolution (SR) reconstruction method by using information fusion and signal processing technology [11] is usually used to obtain HR images higher than the imaging system, which is a commonly used technique in computational imaging processing and practical engineering applications. It is also a hot research issue of underlying computer vision.

SR refers to the technology of constructing HR images from one or more existing LR images in the same scene by using image processing and machine learning technology [12]. Hyperspectral image SR research based on deep learning and sample feature fusion has attracted extensive attention in recent years [13–15]. These methods can effectively improve image spatial resolution while improving spectral resolution by modeling the imaging degradation process and adopting supervised or semi-supervised learning.

Recently, the image SR reconstruction methods based on deep learning have shown great performance [16–24]. Dong et al. [25] proposed a convolutional neural network for super-resolution (SRCNN), which has an end-to-end network architecture, but its network level is too shallow to fully extract image features. Dong et al. [26] further proposed a faster SRCNN (FSRCNN) which selects a smaller size filter compared to SRCNN and introduces a deconvolution layer. Hu et al. [27] proposed that the channel attention block can improve the resolution ability of the network. Lim et al. [28] designed an enhanced deep residual network EDSR by using a simplified residual block. Since then, CNN-based network models have continued to emerge, such as residual dense network RDN [29] and residual feature aggregation network RFANet [30]. Although the image reconstruction performance of the above methods is promising, most network models tend to adopt a standard framework, which is a simple stack of convolution layers, and each layer selects a specific size of convolution kernel. However, the network model parameters greatly increase, occupying more memory and increasing the computational burden. In addition, these methods rarely consider image multi-scale hierarchical features, such as multi-scale features of local spatial information and internal features. The compactness and light weight of the network are worth exploring.

The reconstruction of polarization computational imaging is obtained via accurate analysis. At the same time, the intensity images and polarization parameter images together constitute the observation matrix of object detection. How to make more effective use of the redundant information of the observation matrix in the reconstruction process is a problem to be fully considered in polarization SR.

The above methods applied to the polarization images SR cannot fully exploit the imaging characteristics and prior information, resulting in the limited performance of the model. The input and features of low-resolution polarization images contain rich low-frequency information, which is treated equally in different channels, fails to distinguish and learn across feature channels, and hinders the ability of deep network representation. The quality of image SR is restricted due to the use of the same weights on all paths based on CNNs transmission. To this end, we propose a MSAWN for polarization computational imaging SR which is an end-to-end learning architecture, and design multi-scale local fusion groups (MLFGs) composed of several multi-scale attention blocks (MSABs) to fully analyze the shallow features to the deep features. In each MSAB, the network structure mainly uses parallel dilated convolutions with different expansion rates to capture multi-scale features. In the structure, the multi-scale features are further enhanced by the spatial and channel attention mechanism. Here, a spatial pyramid structure is designed to combine relevant context, average information and depth features of the local fields. Inspired by the above methods, we introduce the adaptive weighted unit (AWU) [31] composed of convolution layers from different receptive fields, which can adaptively remove redundant scale branches to reduce parameters and improve training stability. MSAWN adopts the AWU in the upsampling reconstruction module which gets a parameter that can be learned. Finally, we design a novel reconstruction layer, which utilizes dual paths, as shown in Fig. 1.

The contributions of our work are summarized as follows.

1. We use a limited amount of memory and computational power even with multi-scale and multi-level polarization information. Among the SR methods with similar performance, MSAWN is only about 4.6% of the parameter amount of EDSR, as shown in Table 5.
2. A spatial pyramid structure based on space-channel attention is designed to effectively adjust the feature weight of polarization information. The polarization feature map after attention mechanism contains more negative values, which can well suppress the smooth region in the input polarization image.
(3) Particularly, we design an innovative reconstruction layer. At the end of the network, the features of different paths are fused through a parameter sharing strategy to reconstruct and generate high-resolution polarization images.

(4) The experiments in this paper are mainly based on subjective visual evaluation, objective index evaluation and comparison with the imaging effect of polarization imaging system in Sect. 4.4, which provide a data support for the theoretical research and the system correction of polarization cameras.

The rest of this paper is organized as follows. Section 2 introduces related work. Section 3 illustrates the details of proposed method. The experimental results and analysis are displayed in Sect. 4. The conclusion and future work are presented in Sect. 5.

### 2 Background

#### 2.1 Deep CNNs based network for SR

Nowadays, SR methods are mainly divided into three categories: reconstruction-based methods [32], interpolation-based methods [33], and deep learning-based methods [34]. SR based on deep learning has gradually been a very popular method.

Dong et al. [25] proposed SRCNN which used deep learning for image SR firstly. Since then, SR networks based on deep learning have continuously appeared and improved. Efficient sub-pixel CNN (ESPCNN) [35] and fast super-resolution by CNN (FSRCNN) [26] algorithms directly put LR images into the network to extract features, and learn the HR feature map. Based on residual learning, Kim et al. proposed deeper networks VDSR [23] and DRCN [36] with 20 layers. Tai et al. [37] combined recursive with residual network and proposed a deep recursive residual network DRRN to improve DRCN. Lai et al. [38] designed a laplacian pyramid network LapSRN to gradually increase the image size. Han et al. found that DRCN and DRRN can be uniformly equivalent to a single-state RNN network that only transmits LR spatial information to the HR space. A dual-state recurrent network DSRN [39] is proposed to utilize LR and HR spatial characteristics.

However, as the depth of the network increases, these methods still require excessive memory consumption. Ahn et al. [40] designed a cascaded residual network (CARN) to realize efficient performance and lightweight of the network. Hui et al. [41] proposed an information distillation network to achieve efficient and fast reconstruction. Neural architecture search (NAS) [42] is an emerging method for automatic network design. Chu et al. proposed the NAS-based SR networks More MNA-S [43] and FALSR [44] to achieve a fast, accurate and lightweight model for SR. Due to the limitations of search space and policies in NAS, NAS-based networks require larger computational memory.

#### 2.2 Multi-scale structure based network for SR

To perceive and extract the multi-scale features of objects with different sizes in images, multi-scale network structure has been widely explored and applied in various visual tasks to improve performance. Nowadays, various feature extraction modules have been designed. The main idea of the network is how to explore the optimal local sparse structure based on CNNs.

Christian et al. [45] proposed a deep CNN which is named Inception. To achieve multi-scale receptive field, Inception networks parallelly connect convolution kernels of different scales.

---

**Fig. 1** Network architecture of the proposed MSAWN
sizes in one single layer. Gao et al. [46] proposed a new multi-scale backbone architecture, which is used in various advanced visual tasks. Wang et al. [47] designed generative multi-column convolutional neural networks, which can capture features from different receptive fields, relevant context or phases. Similarly, SinGAN [48] introduced a multi-scale patch discriminator and designed a multi-scale generator that produces images with different resolutions from coarse to fine. Multi-scale structural design also promotes the development of SR. LapSRN [36] restores the HR image by residual learning and stepwise magnification, which adopts multi-scale generation and supervision. Each amplifying module completes double upsampling operations. These modules are of the same structure, share parameters and are cascaded to complete stepwise magnification. Li et al. [49] proposed a multi-scale residual network MSRN for image SR, which utilizes multi-scale hierarchical features to enlarge images at any scale. However, MSRN network extracted features are simply spliced together, which fail to analyze multi-scale context information. Moreover, more problems occur in the training process by blindly increasing the network depth. The parameter size of MSRN is about 6M, which takes up more computing memory and running time.

Based on previous researches, we propose the MLFG, where a spatial pyramid structure is designed to combine relevant context, average information and depth features of the local fields. The multi-scale features extracted in parallel are further enhanced by the spatial and channel attention mechanism. We will show more details in Sect. 3.

3 Methodology

In this section, we will describe the proposed MSAWN in detail. Firstly, an overview of the MSAWN structure is given, in which the process of generating HR from LR Polarization image reconstruction is introduced briefly. Then, MLFG is presented to exploit multi-scale polarization information and polarization multi-level features, in which MLFGs with spatial pyramid structure based space-channel attention will be described. Moreover, the spatial and channel attention mechanism will be elaborated in MSAB, in which channel attention, enhanced spatial attention (ESA) and parallel dilated convolutions are used to enhance representation ability. Besides, AWU is displayed to reduce model parameters in the proposed network. Finally, the loss function is defined to achieve better training.

3.1 Network architecture

As shown in Figs. 1 and 2, our MSAWN mainly includes three parts: shallow feature extraction module (SFEM), nonlinear mapping module (NMM) and adaptive reconstruction module (ARM). Let \( I_{LR} \) and \( I_{SR} \) denote the input and output polarization images of MSAWN, respectively. The shallow feature \( F_0 \) in LR image are extracted by a single \( 3 \times 3 \) convolution layer, which can be formulated as follows:

\[
F_0 = f_0(I_{LR})
\]

where \( f_0(\cdot) \) denotes the shallow feature extraction function of MSAWN. The network sends the extracted feature \( F_0 \) to the backbone module for multi-scale deep feature learning. The main module consists of a long skip connection and multi-scale local fusion groups with several multi-scale attention blocks. This multi-scale feature extraction process can be represented as:

\[
F_1 = f_{msa,n}(\ldots (f_{msa,0}(F_0)) \ldots )
\]

where \( F_1 \) denotes the output feature map of the backbone module, \( f_{msa}(\cdot) \) represents the multi-scale feature extraction function. And then, the multi-scale features extracted in parallel are concatenated as \( F_g = [f_{msa,0}(F_0), f_{msa,1}(F_0), \ldots, F_1] \).

\[
F_2 = H(F_g) \oplus F_0
\]

where \( F_2 \) is the output feature of MLFG, \( H(\cdot) \) denotes the feature extraction function of a \( 3 \times 3 \) convolution layer, and \( \oplus \) means element-wise sum.

The output \( F_3 \) is then fed to the AWU module. Particularly, we adopt a novel reconstruction layer by using sub-pixel convolution, which introduces dual paths. Finally, a \( 3 \times 3 \) convolution layer is used to reconstruct the enlarged feature map to achieve SR. The reconstruction process is as follows:

---

![Fig. 2 Components of the proposed MLFG](image-url)
where \( f_{awu}(\cdot) \) represents the function of the adaptive weighted upscale module, \( \alpha_m \) is an adaptive weight for the \( m \)th scale branch of the adaptive weighted unit, \( f_{up}(\cdot) \) is the function to upsample the features to the object size, and \( f(\cdot) \) denotes the function of MSAWN. The definitions of \( H(\cdot) \) and \( \oplus \) are the same as those in Eq. (3).

### 3.2 Multi-scale attention block

As shown in Fig. 2, MLFG contains several MSABs to fuse multi-scale polarization information and multi-level polarization features gradually. In image reconstruction tasks, MSAB can generate some useful hierarchical features, which are concentrated in different aspects of the original image. But the feature of the first MSAB must go through a long path and repeat addition and convolution operations to reach the last module. To avoid losing some polarization information, skip connections are introduced to improve the stability of deep network training and achieves better performance.

In Fig. 3, a spatial pyramid structure in each MSAB is presented to analyze and extract polarization features from different receptive fields. Here, the spatial and channel attention mechanism [15, 50–52] are introduced to adjust the feature weight of polarization information. To explore the local spatial components of different receptive fields, multi-scale polarization features are extracted by dilated convolutions of different dilation rates (DiRate). Moreover, four dilated convolutions with dilation rates of 1, 2, 4, and 8 are used in the structure to calculate the multi-scale context, which can be formulated as follows:

\[
F_{r,b} = f'_{lock}(F_{r,b-1})
\]

where \( F_{r,b-1} \) denotes the input feature map of MSAB, in which \( r \) is the number of MLFG and \( b-1 \) is the number of MSAB. \( f'_{lock}(\cdot) \) represents the function of \( k \times k \) dilated convolution, in which \( r \) is different dilation rate. \( F_{r,b} \) is the output feature map of dilated convolution layer.

To maximize the effectiveness of the parallel dilated convolution layer, we use the network in conjunction with ESA. This structure concentrates multi-scale features on key spatial content to obtain more representative features. The process with ESA can be formulated as follows:

\[
F_{c,b} = F_{r,b} \odot \delta(F_{r,b})
\]

where \( \delta(\cdot) \) represents the function of the spatial attention mechanism, \( \odot \) means element-wise multiplication, and \( F_{c,b} \) is the output feature map of spatial attention block. And then, the multi-scale features extracted in parallel are concatenated as \( F_{g,b} = \begin{bmatrix} F^1_{s,b}, F^2_{s,b}, F^4_{s,b}, F^8_{s,b}, F^r_{s,b-1} \end{bmatrix} \).

To focus the network on more polarization informative features, the correlation between feature channels can be used. The output \( F_{g,b} \) is then fed to the channel attention block. The output feature \( F_{r,b} \) with channel attention can be represented as:

![Fig. 3 Network structure of spatial pyramid](image-url)
where $\eta_c(\cdot)$ denotes the function of the channel attention mechanism.

Finally, a $3 \times 3$ convolution layer is used to reduce channel dimensions and the weighted multi-scale features are further fully integrated. The output feature is formulated as

$$F_{t,b} = H(F_{c,b})$$

where $H(\cdot)$ denotes the feature extraction function of a $3 \times 3$ convolution layer.

### 3.3 Enhanced spatial attention

Given an object feature and a set of key features, the attention function measures the correlation between the object feature and the pairing of key features to derive attention weights and aggregate the key content adaptively. To enable the model to focus on key content from different representation subspaces and different locations, the output of multiple attention functions is linearly aggregated into learnable weights. To complete well the task of SR, the ESA block with a large receptive field is designed, which is light enough to be inserted into each MSAB. As shown in Fig. 4, the input feature channel dimension is reduced by a $1 \times 1$ convolution layer. Meanwhile, a skip connection is used to directly connect the input feature before spatial dimension reduction to the end of ESA block. Then, a $3 \times 3$ convolution layer with stride 2 is introduced, which can expand the receptive field. The max-pooling layer is used to transform the global spatial information of the input feature into spatial descriptors. To further expand the receptive field, we introduce a $5 \times 5$ convolution group with stride 3. After that, the sub-pixel convolution layer is used to restore the spatial dimension and the channel dimension is restored by a $1 \times 1$ convolution layer. Finally, the sigmoid function is introduced as a simple gating mechanism. The feature representation of the spatial attention mechanism is obtained by correspondingly multiplying the original input feature and the spatial scaling factor.

### 3.4 Channel attention

Since the input features of LR images contain rich low-frequency information that is equally treated in different channels, channel attention mechanism is introduced into the SR network. Moreover, most previous SR network models lack the ability to distinguish learning across feature channels, which hinders the representation ability of networks. Inspired by channel attention mechanism, MSAB with channel attention is designed to extract more informative features. As presented in Fig. 5, the global average pooling layer is introduced to transform the channel global spatial information of the input features into channel descriptors. Let the input be $H \times W \times C$ feature maps, which first pass a global average pooling layer to get a $1 \times 1 \times 1 \times C$ channel description. Then, the weight coefficient of each channel is obtained by a downsampling layer and an upsampling layer. The weight coefficient and the original feature can be multiplied to obtain the new feature after scaling. The whole process is actually a weighted distribution of the features of different channels. Among them, the downsampling and upsampling layers are implemented by $1 \times 1$ convolutions. The number of channels in the downsampling layer is reduced by $r$ times and the activation function is ReLU. The activation function of the upsampling layer is sigmoid. The process with channel attention can be formulated as follows:

$F = \varepsilon_{gp}(F_{in})$

$F_{out} = f_u(w_{up}(w_d)) \cdot F_{in}$

![Fig. 4 The structure of enhanced spatial attention block (ESA)](image)

![Fig. 5 Network structure of channel attention mechanism](image)
where $F_{in}$ and $F_{out}$ denote the input and output features of the network, $\epsilon_{gp}(\cdot)$ represents the function of the global average pooling layer, $F$ denotes the output feature of the global average pooling layer, $w_q$ and $w_{up}$ are the weights of the convolution layer, and $f_\alpha(\cdot)$ and $\beta(\cdot)$ are the sigmoid function and ReLU function.

### 3.5 Adaptive weighted unit

The network model is implemented by increasing the depth, which greatly increases the parameters and increases the computational burden. The AWU module aims to adaptively remove redundant scaling branches to reduce parameters. AWU removes the fully connected layer of generated parameters and sets a parameter function that can be learned. The network exploits this function to replace the global average pooling and fully connected layer for learning parameters, thus reducing numerous parameters. As shown in Fig. 1, the input features first pass through a convolution group of convolution kernels with multiple sizes of 3, 5, 7, and 9 to extract feature information from different receptive fields and adaptively learn a parameter for each receptive field. Then, the obtained features are subjected to sub-pixel convolution upsampling and multi-scale feature fusion (the feature maps extracted from each receptive field are added element by element). Instead of merely adopting most SR methods of deconvolution or sub-pixel convolution single-path upsampling, the upscale module introduces dual paths, which combine with LR image features using sub-pixel convolution.

### 3.6 Loss function

The common loss functions used in the previous SR methods include $L_1$, $L_2$, perceptual loss, and adversarial loss. We choose the $L_1$ loss function to optimize the MSAWN model. Let $x$ be the input LR image and $\theta$ be the network parameter set to be optimized. The goal is to learn a mapping function $f(\cdot)$ and use the function $\hat{y} = f(x; \theta)$ to generate HR images. Assuming that a set of training sets is given, the HR image is first scaled by bicubic interpolation downsampling, and the mapping relationship of each data set of HR-LR pairing with different downsampling scales is studied. Instead of minimizing the mean squared error (MSE) between HR and $\hat{y}$, we adopt a stable $L_1$ loss function to process the loss value. The goal of training MSAWN is to minimize the $L_1$ loss function, which can be expressed as:

$$L(\theta) = \frac{1}{N} \sum_{i=1}^{N} \left\| f(\mathbf{I}_{LR}^i) - \mathbf{I}_{HR}^i \right\|_1$$

(11)

where $f(\cdot)$ denotes the function of our SR network, and $\mathbf{I}_{LR}^i$ and $\mathbf{I}_{HR}^i$ respectively represent the $i$th LR image in the training data and its corresponding HR image. We utilize stochastic gradient descent to optimize the loss function.

### 4 Experiments

In this section, we first introduce the datasets and metrics in detail, and then the implementation details are described. Next, the analyses of each network component is presented to explore its role. Then, we conduct comparative evaluations in terms of visual quality, time complexity, and quantitative evaluation. Finally, the SR reconstructed image is compared with the real image collected by the actual hardware to provide data reference for system calibration and correction.

#### 4.1 Datasets and metrics

As shown in Fig. 6, the polarization image is developed by a set of spatially modulated infrared and visible light dual-channel polarization cameras based on Savart polarizer, which can quickly measure and collect the complete polarization state of the object under a given spectrum. The imaging system adopts the spatial modulation principle of Stokes vector [9], and modulates four Stokes vectors (i.e. $S_0\sim S_3$) in the same image at the same time. The modulation information containing four Stokes vectors of the object can be obtained by one acquisition in the system. Based on this, multiple polarization parameter images can be resolved. We employ self-built polarization image data sets (i.e. 800 images) for training experiments. During testing, polarization images (i.e. $640 \times 512$) with different background are used, which are enough to validate the model. Following previous works, PSNR and SSIM evaluated on the Y channel (i.e. luminance component) of the transformed YCbCr space [26] are used as evaluation criteria. Furthermore, information entropy and definition indicators are used to evaluate image quality without reference.

![Fig. 6 Spectral polarization camera. a Experimental camera. b Experimental scenario](image-url)
4.2 Implementation details

During the training, the training images are randomly rotated $90^\circ$, $180^\circ$, $270^\circ$ and flipped horizontally for data augmentation. We use the bicubic degradation model to down-sample the high-definition training data set collected by the polarization camera with different multiples and add blur to obtain the corresponding LR images. The number of MLFGs is set to 4, and the number of MSABs is set to 4 by default in our network structure. Moreover, we set the number of input channels as 32. In each batch of training, LR samples with a size of $48 \times 48$ are extracted as input, and the network is iteratively trained for 1000 epochs. The batch-size is set to 16. For AWU module, a convolution group of convolution kernels with multiple sizes of 3, 5, 7, and 9 is set, where the initial weight for each scale branch is 0.25. Our model is trained with the Adam optimizer [54] with an initial learning rate of $10^{-3}$. After each 200 epoch training (i.e. 2e5 iterations), the learning rate of back-propagation iteration decreases by half. The MSAWN is implemented with the Pytorch framework on a Nvidia Quadro RTX 6000 GPU.

4.3 Ablation study

In this section, we discuss the roles of the different components of MSAWN via ablation experiments. As shown in Figs. 7, 8, the ablation experiments achieve on bridge and building polarization images with different scales (including $\times 3$, $\times 4$).

(1) Multi-scale attention block (MSAB) module: In each MSAB, a spatial pyramid structure based on space and channel attention is designed to extract features from different receptive fields. To demonstrate the value of channel attention unit and enhanced spatial attention block, we remove them from the MSAB framework respectively to reconstruct other parts remained. The quantitative results achieved on building polarization images are shown in Table 1. The results show that our MSAB module presents superior reconstruction indicators to other networks. Compared with the CA-only network, the PSNR of ESA-only network is slightly higher. This explicitly validates that SR performance are further enhanced by the spatial and channel attention mechanism of our model from 34.42 to 34.72 dB for $\times 4$ SR.

(2) Adaptive weighted unit (AWU) module: Most of SR methods based on deep learning occupy excessive amounts of computational and memory consumption, which are difficult to apply to real applications. The AWU module aims to adaptively remove redundant scaling branches to reduce parameters. The MSAWN can be implemented with the Pytorch framework on a Nvidia GeForce GTX 1650Ti GPU (4 GB memory) without demanding amounts of memory consumption. To illustrate the effectiveness of the AWU module, we remove it from our model and reconstruct other parts remained. As shown in Table 2, we can clearly see that our model with the AWU module achieve better performance and the PSNR is significantly improved to be 34.72 dB for $\times 4$ SR. This indicates that our model

| CA | ESA | PSNR (dB) | SSIM |
|----|-----|----------|------|
| $\times$ | $\times$ | 34.42 | 0.9251 |
| $\times$ | ✓ | 34.59 | 0.9267 |
| ✓ | $\times$ | 34.45 | 0.9257 |
| ✓ | ✓ | 34.72 | 0.9281 |
benefits from the AWU module which achieves superior SR reconstruction performance.

(3) Upscale module: To fully exploit the information in LR image space, we adopt a novel reconstruction layer, which introduces dual paths, combined with LR image features using sub-pixel convolution. To explore the efficiency of upscale module, we retrain the network with a single sub-pixel convolution layer to upsample the features to the object size. Compared to our proposed MSAWN, the PSNR of the network without a novel reconstruction layer drops from 34.72 to 34.56 dB for ×4 SR in Table 3. This demonstrates that a novel reconstruction layer enables our model to improve SR performance. Thus, such that the information in the LR image space is effectively extracted, which promotes superior reconstruction quality.

(4) Network depth: SR reconstruction performance can be effectively improved by increasing the depth of the network. To gain excellent results, we change the number of multi-scale local fusion groups in our work. However, as the depth of the network increases, the model becomes more difficult to train, occupying more memory and aggravating the computational burden. To explore the impact of multi-scale local fusion groups, we train our model with different number of MLFGs (including 2, 3, 4, 5). The quantitative results achieved on bridge polarization images are shown in Table 4. It can be seen that the PSNR value increases gradually from 44.80 to 45.03 dB for ×3 SR with the network deepening. However, when the number of MLFGs is 4, we get better PSNR value. While weighing SR performance and model parameters, we select the network with four MLFGs.

### 4.4 Results

#### 4.4.1 Comparison against other methods

As shown in Fig. 9, the comparative experiments achieve on bridge, building and road polarization images with different scales (including ×2, ×3, ×4). To assess the effectiveness of the MSAWN in polarization images SR, our model is compared with some state-of-the-art methods, including Bicubic, SRCNN [25], FSRCNN [26], MSRN [49], EDSR [28], LW-AWSRN [31]. Quantitative comparisons on different polarization images are presented in Table 5. Furthermore, the qualitative analysis is illustrated by visual graphics. The running time and parameters of the above models are also provided for more intuitive comparison. Quantitative and qualitative evaluations are further described as follows. From Table 5, we can see that the proposed MSAWN yields similar or better performance compared to the other approaches. It is noted that MSAWN obtains superior performance for ×2 SR and our results are almost close to the best results for ×3 and ×4 SR on different polarization images. Our model parameters are only about 1/3 of the parameters of MSRN. In particular, our model exceeds existing lightweight networks: LW-AWSRN-S and FSRCNN. Our experiment has achieved almost the equivalent results as that of EDSR, but our model is only about 4.6% of the parameters of EDSR. Overall, the proposed method has significantly fewer parameters than MSRN while achieving higher performance, which provides a better trade-off between parameters and effects. Considering the computation cost, the MSAWN is more suitable for real applications.

For qualitative analysis, three detailed, textured images selected from the polarization images in Fig. 9 above are tested for comparison. The visual comparison results of different methods are presented in Figs. 10, 11, 12. It is observed that the SR reconstruction image quality of the proposed network is superior to other methods in terms of sharpness, brightness, or texture.

In terms of running time, the above methods are utilized to evaluate the running time of bridge images for ×3 SR. Both of the running time and complexity are expressed through Fig. 13. As shown in Fig. 13, we can see that the MSAWN has faster execution to reconstruct the image compared to the MSRN and EDSR methods, which occupies less memory consumption. It is notable that the proposed method sacrifices a small amount of time, improves the accuracy, and the time-consuming is about 0.5 s, which has better real-time performance.

Overall, the proposed method is superior to other methods through the above quantitative and qualitative analysis.

#### Table 3 SR results of our upscale module on building polarization images at four scale factor

| Upscale module | PSNR (dB) | SSIM |
|----------------|----------|------|
| ×              | 34.56    | 0.9262 |
| ✓              | 34.72    | 0.9281 |

#### Table 2 Effects of the AWU on building polarization images for ×4 SR

| AWU | PSNR (dB) | SSIM |
|-----|----------|------|
| ×   | 34.64    | 0.9269 |
| ✓   | 34.72    | 0.9281 |
This indicates that the MSAWN is more effective in polarization image SR.

4.4.2 Full polarization image analysis

The full-polarization imaging system acquires the spectral characteristics, specific spatial characteristics and the relevant intrinsic polarization information reflecting the target material through modulation and analysis. The final result of the full polarization computational imaging system is to resolve the full polarization parameters. It reflects the scene and detection target information from different angles, and the full polarization parameters of the target realize synchronous detection. To further reflect the effectiveness of our method and calibrate and correct the system, we select the high-definition monument image to obtain the test sample of this experiment through modulation and analysis. Then, we randomly selected six detailed and well-textured analytic images of monuments under earth observation to form the test image set. As shown in Fig. 14, they are an $E_x$ image, an $E_y$ image, an I image, a Q image, a U image, and a V image, respectively. In this experiment, PSNR and SSIM are used to evaluate the reconstruction effect of the images.

From the experimental data in Table 6, it can be seen that the analytical results of different polarization parameters have different test performances. However, our method shows good results in the analytical results of most of the polarization parameters. The SSIM evaluation index can basically maintain about 0.95 in the two times and three times SR tests, which fully verifies the effectiveness of our method in reconstructing the analytical images of most polarization parameters. The PSNR index shows a downward trend as

![Different polarization images](image-url)

**Table 5** Quantitative results on different polarization images (PSNR (dB)/SSIM)

| Scale | Method      | Params | Bridge image | Building image | Ship image | Road image |
|-------|-------------|--------|--------------|----------------|------------|------------|
|       | Bicubic     | –      | 42.49/0.9671 | 42.42/0.9693  | 44.85/0.9875 | 42.28/0.9596 |
| 2×2   | SRCNN 57K   | 43.82/0.9727 | 43.38/0.9749 | 45.86/0.9886 | 42.92/0.9657 |
|       | FSRCNN 12K | 43.85/0.9729 | 43.39/0.9749 | 44.47/0.9865 | 42.88/0.9654 |
| ×3    | LW-AWSRN-S 397K | 45.08/0.9775 | 44.60/0.9786 | 48.25/0.9916 | 44.07/0.9710 |
|       | MSRN 5.93M  | 45.12/0.9775 | 44.63/0.9787 | 48.39/0.9917 | 44.11/0.9711 |
|       | EDSR 40.7M  | 45.13/0.9775 | 44.65/0.9787 | 48.43/0.9917 | 44.13/0.9712 |
|       | MSAWN (ours) 1.95M | 45.12/0.9776 | 44.65/0.9787 | 48.40/0.9917 | 44.13/0.9712 |
|       | Bicubic     | –      | 39.25/0.9450 | 39.32/0.9457 | 40.34/0.9690 | 40.07/0.9356 |
| 2×2   | SRCNN 57K   | 41.55/0.9576 | 40.97/0.9583 | 42.34/0.9746 | 40.99/0.9464 |
|       | FSRCNN 12K | 41.82/0.9589 | 41.13/0.9591 | 41.79/0.9719 | 41.02/0.9465 |
| ×3    | LW-AWSRN-S 477K | 43.07/0.9670 | 42.63/0.9672 | 44.66/0.9809 | 42.33/0.9567 |
|       | MSRN 6.11M  | 43.37/0.9670 | 42.77/0.9678 | 45.13/0.9817 | 42.39/0.9571 |
|       | EDSR 43.6M  | 43.52/0.9685 | 42.86/0.9682 | 45.39/0.9821 | 42.41/0.9572 |
|       | MSAWN (ours) 2.03M | 43.48/0.9679 | 42.86/0.9677 | 45.33/0.9815 | 42.41/0.9572 |
|       | Bicubic     | –      | 36.78/0.9199 | 36.48/0.9121 | 37.14/0.9513 | 38.10/0.9080 |
| 2×2   | SRCNN 57K   | 38.45/0.9328 | 38.05/0.9303 | 39.09/0.9567 | 38.89/0.9194 |
|       | FSRCNN 12K | 39.36/0.9401 | 38.42/0.9328 | 39.28/0.9541 | 38.96/0.9193 |
| ×4    | LW-AWSRN-S 588K | 41.24/0.9563 | 40.29/0.9498 | 42.33/0.9715 | 40.58/0.9388 |
|       | MSRN 6.08M  | 41.41/0.9571 | 40.40/0.9509 | 42.62/0.9722 | 40.68/0.9397 |
|       | EDSR 43M    | 42.19/0.9609 | 40.70/0.9531 | 42.97/0.9731 | 40.70/0.9398 |
|       | MSAWN (ours) 2.14M | 42.12/0.9589 | 40.70/0.9515 | 42.89/0.9719 | 40.71/0.9398 |
the scale of super-resolution increases, but our method still achieves good results on an I image, a Q image, a U image, and a V image. In general, the above experimental data fully demonstrate the superiority of our method in resolving images with different polarization parameters.

4.4.3 System calibration experiments

The spectral polarization camera we used consists of two channels, one is a visible near-infrared spectral polarization camera and the other is a short-wave infrared spectral polarization camera. The polarization imaging system can realize the adjustment of the imaging spatial resolution according to the detection object and the requirements of image processing. For visible polarized images with resolution of 2048 × 2048 and short-wave infrared polarized images with resolution of 640 × 512, we can separately collect HR images and LR (i.e. 1024 × 1024, 320 × 256) images by setting the system. To compare the imaging effect of our method and the imaging system, we perform SR reconstruction on the LR image. Then the SR image is evaluated with the HR image. Calibration experiments use information entropy and definition to evaluate image quality without reference. As explained in Table 7, we can see that in terms of definition, the SR reconstruction quality is
generally slightly lower than the system HR image quality, but higher than the system LR image.

For the information entropy, the SR reconstruction result is basically consistent with the system HR image effect, but it is improved compared to the LR image. Figure 15 exhibits that HR and LR images collected by the system are compared with the SR reconstructed image. It can be observed that the LR image faces problems such as blurring of bump stripes, obvious blurring of the background, and unclear texture. The image with bump stripes reconstructed by our method has clear texture details, uniform brightness, and is close to the HR image. Compared with the LR image, the reconstructed quality is significantly improved. As shown in Fig. 16, The HR image curve tends to be flat, and the SR image curve shows a downward trend, but it is close to the HR image curve in terms of information entropy. For the definition, the curves of HR and SR images tend to be gentle and basically flat, with a small spacing. The above experimental data fully verify the effectiveness of our method, and the reconstructed image is close to the effect of system imaging.
5 Conclusion

In this paper, the MSAWN is proposed for polarization computational imaging SR with limited computational resources. We design a spatial pyramid structure with the spatial and channel attention mechanism to comprehensively incorporate multilevel information. Based on the spatial pyramid structure, we propose the MLFG to extract features from different receptive fields. In addition, we introduce an AWU to adaptively remove redundant scaling branches to reduce parameters. Particularly, we design a novel reconstruction layer with dual paths to upscale the features. The extensive experiments...
illustrate that our method achieves better reconstruction accuracy and visual improvement effects. The proposed method also provides a data support for the theoretical research on polarization images SR and the system correction of polarization cameras. The future work is to reduce the amount of parameters in the network while optimizing training efficiency.

Acknowledgements This work is supported by the National Natural Science Foundation of China under Grant No. 61906118 and Natural Science Foundation of Anhui Province under Grant Nos. 1908085MF208 and 2108085MF230; Key Project of Natural Science Research in Colleges and Universities of Anhui Province (No. KJ2019A0906), and the Major Special Science and Technology Project of Anhui Province (No.202003A06020016).

Data availability The datasets analysed during the current study are available from the corresponding author on reasonable request.

Declarations

Conflict of interest The authors declare no conflicts of interest.

References

1. V. Thilak, D.G. Voelz, C.D. Creusere, Polarization-based index of refraction and reflection angle estimation for remote sensing applications. Appl. Opt. 46(30), 7527–7536 (2007)
2. F. Cremer, W. De Jong, K. Schutte, Infrared polarization measurements and modelling applied to surface laid anti-personnel landmines. Opt. Eng. 41(5), 1021–1032 (2002)
3. N. Yadav, W. Wang, K. Ullah, X. Liu, Polarization parametric indirect microscopic imaging for patterned device line edge inspection. Appl. Phys. B 124(8), 1–8 (2018)
4. J. Wang, L. Dong, H. Chen, S. Huang, Birefringence measurement of biological tissue based on polarization-sensitive digital holographic microscopy. Appl. Phys. B 124(12), 1–9 (2018)
5. S. Wen, Y. Zheng, F. Lu, Q. Zhao, Convolutional deconvolution network for joint chromatic and polarimetric imagery. Opt. Lett. 44(22), 5646–5649 (2019)
6. W.A. Woźniak, P. Kurzynowski, S. Drobczyński, Adjustment method of an imaging stokes polarimeter based on liquid crystal variable retarders. Appl. Opt. 50(2), 203–212 (2011)
7. J. Zhang, C. Yuan, G. Huang, Y. Zhao, W. Ren, Q. Cao, J. Li, M. Jin, Acquisition of a full-resolution image and aliasing reduction for a spatially modulated imaging polarimeter with two snapshots. Appl. Opt. 57(10), 2376–2382 (2018)
8. L. Tong, X. Huang, P. Wang, L. Ye, M. Peng, L. An, Q. Sun, Y. Zhang, G. Yang, Z. Li et al., Stable mid-infrared polarization imaging based on quasi-2d tellurium at room temperature. Nat. Commun. 11(1), 1–10 (2020)
9. V. Durán, P. Clemente, M. Fernández-Alonso, E. Tajahuerce, J. Lancis, Single-pixel polarimetric imaging. Opt. Lett. 37(5), 824–826 (2012)
10. W. Treimer, R. Köhler, Determination of the spatial resolution in the case of imaging magnetic fields by polarized neutrons. Appl. Sci. 11(15), 6973 (2021)
11. K. Jiang, Z. Wang, P. Yi, J. Jiang, A progressively enhanced network for video satellite imagery superresolution. IEEE Signal Process. Lett. 25(11), 1630–1634 (2018)
12. K. Zhang, W. Zuo, L. Zhang, Deep plug-and-play super-resolution for arbitrary blur kernels. In: Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., pp. 1671–1681 (2019)
13. Q. Xie, M. Zhou, Q. Zhao, D. Meng, W. Zuo, Z. Xu, Multispectral and hyperspectral image fusion by ms/hs fusion net. In: Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., pp. 1585–1594 (2019)
14. W. Wang, W. Zeng, Y. Huang, X. Ding, J. Paisley, Deep blind hyperspectral image fusion. In: Proc. IEEE Int. Conf. Comput. Vis., pp. 4150–4159 (2019)
15. Y. Hu, J. Li, Y. Huang, X. Gao, Channel-wise and spatial feature modulation network for single image super-resolution. IEEE Trans. Circuits Syst. Video Technol. 30(11), 3911–3927 (2019)
36. J. Kim, J.K. Lee, K.M. Lee, Deeply-recursive convolutional network
35. W. Shi, J. Caballero, F. Huszár, J. Totz, A.P. Aitken, R. Bishop, D.
34. Q. Yuan, Y. Wei, X. Meng, H. Shen, L. Zhang, A multiscale and
20. Z. Pan, W. Ma, J. Guo, B. Lei, Super-resolution of single remote
19. W. Zhang, Y. Liu, C. Dong, Y. Qiao, Ranksrgan: Generative adversarial networks with ranker for image super-resolution. In: IEEE Trans. Pattern Anal. Mach. Intell., pp. 3096–3105 (2019)
18. D. Yuan, N. Fan, Z. He, Learning target-focusing convolutional regression model for visual object tracking. Knowl. Based Syst. 194, 105526 (2020)
17. Y. LeCun, Y. Bengio, G. Hinton, Deep learning. Nature 521(7553), 436–444 (2015)
16. Z. Wang, Y. Liu, C. Dong, Y. Qiao, Ranksrgan: Generative adversarial networks with ranker for image super-resolution. In: Proc. IEEE Trans. Pattern Anal. Mach. Intell., pp. 3096–3105 (2019)
15. J. Kim, J.K. Lee, K.M. Lee, Deeply-recursive convolutional network
14. W. Shi, J. Caballero, F. Huszár, J. Totz, A.P. Aitken, R. Bishop, D.
13. Q. Yuan, Y. Wei, X. Meng, H. Shen, L. Zhang, A multiscale and
12. Z. Pan, W. Ma, J. Guo, B. Lei, Super-resolution of single remote sensing image based on residual dense backprojection networks. IEEE Trans. Geosci. Remote Sens. 57(10), 7918–7933 (2019)
11. Y. Chen, H. Jiang, C. Li, X. Jia, P. Ghamisi, Deep feature extraction and classification of hyperspectral images based on convolutional neural networks. IEEE Trans. Geosci. Remote Sens. 54(10), 6232–6251 (2016)
10. X. Hu, H. Mu, X. Zhang, Z. Wang, T. Tan, J. Sun, Meta-sr: A magnification-arbitrary network for super-resolution. In: Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., pp. 1575–1584 (2019)
9. J. Kim, J.K. Lee, K.M. Lee, Accurate image super-resolution using very deep convolutional networks. In: Proc. IEEE/ICIP Conf. Pattern Recognit., pp. 1646–1654 (2016)
8. J.M. Haut, R. Fernandez-Beltran, M.E. Paoletti, J. Plaza, A. Plaza, F. Pla, A new deep generative network for unsupervised remote sensing single-image super-resolution. IEEE Trans. Geosci. Remote Sens. 56(11), 6792–6810 (2018)
7. C. Dong, C.C. Loy, K. He, X. Tang, Image super-resolution using deep convolutional networks. IEEE Trans. Pattern Anal. Mach. Intell. 38(2), 295–307 (2015)
6. C. Dong, C.C. Loy, X. Tang, Accelerating the super-resolution convolutional neural network. In: Proc. Comput. Vis.-ECCV, pp. 391–407 (2016)
5. J. Hu, L. Shen, G. Sun, Squeeze-and-excitation networks. In: Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., pp. 7132–7141 (2018)
4. B. Lim, S. Son, H. Kim, S. Nah, K. Mu Lee, Enhanced deep residual networks for single image super-resolution. In: Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops, pp. 136–144 (2017)
3. Y. Zhang, Y. Tian, Y. Kong, B. Zhong, Y. Fu, Residual dense network for image super-resolution. In: Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., pp. 2472–2481 (2018)
2. J. Liu, W. Zhang, Y. Tang, J. Tang, G. Wu, Residual feature aggregation network for image super-resolution. In: Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., pp. 2359–2368 (2020)
1. Z. Li, C. Wang, J. Wang, S. Ying, J. Shi, Lightweight adaptive weighted network for single image super-resolution. Comput. Vis. Image Underst. 211, 103254 (2021)