

Article

A Super-Resolution Reconstruction Network of Space Target Images Based on Dual Regression and Deformable Convolutional Attention Mechanism

Yan Shi ^{1,2}, Chun Jiang ^{1,2}, Changhua Liu ¹, Wenhan Li ^{1,2} and Zhiyong Wu ^{1,*}

¹ Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences, Changchun 130033, China

² University of Chinese Academy of Sciences, Beijing 100049, China

* Correspondence: wuzy@ciomp.ac.cn

Abstract: High-quality space target images are important for space surveillance and space attack defense confrontation. To obtain space target images with higher resolution and sharpness, this paper proposes an image super-resolution reconstruction network based on dual regression and a deformable convolutional attention mechanism (DCAM). Firstly, the mapping space is constrained by dual regression; secondly, deformable convolution is used to expand the perceptual field and extract the high-frequency features of the image; finally, the convolutional attention mechanism is used to calculate the saliency of the channel domain and the spatial domain of the image to enhance the useful features and suppress the useless feature responses. The experimental results show that the method outperforms the comparison algorithm in both objective quality evaluation index and localization accuracy on the space target image dataset compared with the current mainstream image super-resolution algorithms.

Keywords: super-resolution reconstruction; space target; dual regression; deformable convolutional attention mechanism



Citation: Shi, Y.; Jiang, C.; Liu, C.; Li, W.; Wu, Z. A Super-Resolution Reconstruction Network of Space Target Images Based on Dual Regression and Deformable Convolutional Attention Mechanism. *Electronics* **2023**, *12*, 2995. <https://doi.org/10.3390/electronics12132995>

Academic Editor: Silvia Liberata Ullo

Received: 15 May 2023

Revised: 3 July 2023

Accepted: 4 July 2023

Published: 7 July 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Space targets mainly include satellites, spacecraft, various space debris, various space vehicles entering the Earth's outer space, and deep space objects. With the human exploration and utilization of space, the number of spacecraft is increasing, which also leads to the gradual increase in space debris, and the resulting space environmental safety issues are also of increasing concern.

In real-time space situational awareness, a space target detection system with a ground-based large field-of-view optoelectronic telescope plays an important role, but there are still major technical problems that have not been solved. The images taken by the large field of view space target photodetection system contain a large number of stellar targets and space targets, and the severe interference of atmospheric turbulence, the performance of sensor hardware equipment, and the limitation of additional noise make the target images obtained by ground-based photoelectric telescopes appear severely blurred and degraded. The degraded image shows the characteristics of low definition and low signal-to-noise ratio, the image is blurred and the details are indistinguishable; the target information that can be obtained is very limited, and it is difficult to accurately detect and locate the space target, so it is necessary to improve the space target image quality and increase the image resolution effectively.

According to the diffraction limit formula, one of the most effective ways to improve the imaging resolution of a telescope is to increase the aperture of the primary mirror of the telescope. However, increasing the aperture of the primary mirror causes an increase in the weight and volume of the mirror, and the corresponding support structure of the

mirror tends to be complex and large; secondly, the basic requirement of keeping the root mean square error (RMS) of the reflector surface smaller than $\lambda/20$ at large apertures becomes extremely demanding, so the processing difficulty and manufacturing cost increase geometrically with the size of the primary mirror [1,2]. Improvements to imaging systems at the hardware level to improve resolution are limited by many factors, making the super-resolution reconstruction of low-resolution images by algorithms at the software level an important means of improving image resolution.

Image super-resolution (SR) reconstruction is a technique that converts existing low-resolution (LR) images into high-resolution (HR) images through software algorithms using signal processing and image processing methods. SR reconstruction technology can make the image store more information per unit area. Compared with the low-resolution images, high-resolution images can represent richer detail information and have stronger information expression ability. Therefore, SR reconstruction technology can not only improve the display effect of images but also help with the further analysis and processing of images, which is important for the subsequent detection, tracking and localization of space targets.

In recent years, with the rapid development of deep learning, SR reconstruction methods for images based on deep learning have made remarkable progress. In the field of natural image processing, deep learning-based SR algorithms have achieved good reconstruction results on publicly available image datasets. However, they are not widely used in the field of SR reconstruction of space target images. On the one hand, they are limited by the low quality of the space target images themselves, and on the other hand they are also limited by the small number of publicly available training sets of space target images. According to the characteristics of space target images, this paper puts forward an image super-resolution dual regression network based on deep learning, constructs a deep learning training set of space target images, and conducts network training, aiming at realizing the clear reconstruction of space target images, reducing image artifacts, enriching image details and improving positioning accuracy. The specific research contents of this paper are as follows.

1. To recover high-quality space target images with less computational cost, a dual structure for SR reconstruction is used in this paper. Compared with the traditional algorithm, which only has a mapping relationship from the LR image to the HR image, the proposed method in this paper also adds inverse mapping to support SR reconstruction work.
2. The introduction of deformable convolution to expand the perceptual field can adaptively find features that are more useful for the present and extract the high-frequency characteristics of the image.
3. The space target image, as a single-channel image, contains less information than the natural image in terms of both dimensionality and quantity. To address this problem, this paper introduces the attention mechanism. The convolutional attention mechanism was used to compute the saliency of the channel domain and the spatial domain of the image to extract deeper features that are more accurate and effective.

2. Related Work

2.1. Image Super-Resolution Reconstruction

In 1964, Harris [3] studied the physical limit of resolution of optical imaging systems, which laid the mathematical foundation for image super-resolution. In 1984, Tsai et al. [4] obtained an HR image with Fourier transforms domain processing using multiple LR images, which was the first attempt to use software technology for image SR reconstruction. Image SR reconstruction can be divided into interpolation-based, reconstruction-based, and learning-based methods [5], which have been applied to many fields such as medical imaging [6,7], security monitoring [8], and remote sensing imaging [9].

2.1.1. Interpolation-Based Methods

Interpolation-based methods [10,11] are used to perform image SR reconstruction by exploiting the existence of a correlation between neighboring pixels of the original image, which can give better results even when the training samples are insufficient. It can usually be divided into nearest neighbor interpolation [12], bilinear interpolation [13], and bicubic interpolation methods [14]. The basic problem of current interpolation-based image SR methods is that it is difficult to generate new high-frequency information, and although these methods are fast and simple compared with other image SR methods, they still have effects such as jaggedness and blurriness. Therefore, interpolation-based image SR algorithms cannot meet the requirements of most image SR reconstruction applications.

2.1.2. Reconstruction-Based Methods

The reconstruction-based method is the mainstream image SR reconstruction method before the emergence of the learning-based method. Reconstruction-based methods [4,15] usually introduce prior knowledge as constraints in the reconstruction process to improve the details of the reconstructed image, such as in the form of noise perturbation, energy function, etc., or perform iterative computations to approximate the original high-resolution image. Therefore, reconstruction-based methods are usually computationally intensive, difficult to solve, and time-consuming, and cannot meet the task requirements of high-precision image super-resolution.

2.1.3. Learning-Based Methods

To reconstruct super-resolution images with high precision, researchers proposed a learning-based method to learn the mapping relationship between high-resolution and low-resolution image pairs. The learning-based SR technique [16] was first proposed by Freeman, and its basic idea is to train before reconstruction. In the training process, the mapping relationship between LR images and their corresponding HR images is learned by training samples, and then in the reconstruction process the input LR images are used to predict the HR images based on the learned mapping relationship to achieve SR reconstruction.

The learning-based SR reconstruction technique can be summarized in the following three steps: first, the network model from the LR image to the HR image is designed based on the prior knowledge of the image; second, the network is trained to learn the mapping relationship from the LR image to the HR image by relying on the training samples. Finally, the learned model uses the LR image to predict the corresponding HR image to achieve SR reconstruction.

In recent years, deep learning has achieved great success in many image-processing problems. In 2015, Dong et al. [17] first applied convolutional neural networks to SR reconstruction problems and proposed the image SR algorithm SRCNN based on convolutional neural networks. Its emergence solved many bottlenecks in traditional SR techniques and highlighted the excellent performance of convolutional neural networks in SR problems. So far, a large number of excellent SR algorithms based on deep learning have emerged.

Overall, the deep learning-based SR algorithm is the most complex among the three algorithms and also has the best effect. Compared with the traditional SR algorithm, the deep learning-based SR algorithm is more capable of learning the nonlinear mapping relationship between images, can learn higher-level image features, and has a stronger generalization ability, so it is the main research direction in the field of SR reconstruction at present.

2.2. Fundamentals of Space Target Image Detection Technology

2.2.1. Imaging of Space Target

In general, optical telescopes operate in two modes: a “stellar gazing mode”, in which the telescope is constantly adjusted to gaze at a fixed star, so the space target appears as a

bar; and a “target gazing mode”, in which the telescope is constantly readjusted to gaze at a fixed space target, so the space target appears as a point.

In this paper, we focus on the images of space targets acquired in the “stellar gazing mode” of the telescope, as shown in Figure 1, in which the stars appear as dots and their positions are constant between adjacent frames, and the space target is in the shape of a dashed line and the same direction as the motion of the stars.



Figure 1. Image taken by a ground-based telescope in “stellar gazing mode”.

2.2.2. Endpoint Localization Technology of the Space Target

This paper proposes a strip target endpoint detection method based on Harris corner point detection. The Harris corner point detection algorithm [18] is a corner point feature extraction operator proposed by Harris and Stephens in 1988. The basic idea of corner point detection is that a detection window is used to move in any direction on the image, and its analysis is performed by comparing the degree of grayscale change of pixels in the window when sliding, and if there is a slide in any direction that has a large grayscale change, then we can assume that there is a corner point in the window. The autocorrelation function of the image window translation $[u, v]$ that produces the grayscale change is

$$E(u, v) = \sum_{x,y} w(x, y) [I(x + u, y + v) - I(x, y)]^2, \tag{1}$$

where $w(x, y)$ is the window function, $I(x + u, y + u)$ is the image grayscale after translation, and $I(x, y)$ is the image grayscale.

$$\sum_{x,y} [I(x + u, y + v) - I(x, y)]^2, \tag{2}$$

can be expanded and intercepted by Taylor formula, which can be approximated as

$$\sum_{x,y} [I(x + u, y + v) - I(x, y)]^2 \approx [u, v] \left[\sum \begin{bmatrix} I_x^2 & I_x I_y \\ I_x I_y & I_y^2 \end{bmatrix} \right] \begin{bmatrix} u \\ v \end{bmatrix}, \tag{3}$$

then

$$E(u, v) \approx [u, v] \left[\sum_{x,y} w(x, y) \begin{bmatrix} I_x^2 & I_x I_y \\ I_x I_y & I_y^2 \end{bmatrix} \right] \begin{bmatrix} u \\ v \end{bmatrix}, \tag{4}$$

let $M = \sum_{x,y} w(x, y) \begin{bmatrix} I_x^2 & I_x I_y \\ I_x I_y & I_y^2 \end{bmatrix}$, and the response strength of characteristic points is defined as

$$R = \det M - k(trM)^2, \tag{5}$$

and according to experience, $k = 0.04 \sim 0.06$.

For striped targets, the gray gradient changes little when following the target stripe direction, and the gradient changes rapidly in both horizontal and vertical directions at the endpoint position, and there is a significant change in the grayscale within the window,

indicating the presence of corner points. Therefore, this method can be used to locate the endpoints of striped targets. The two endpoints of the localized space target image are shown in Figure 2.

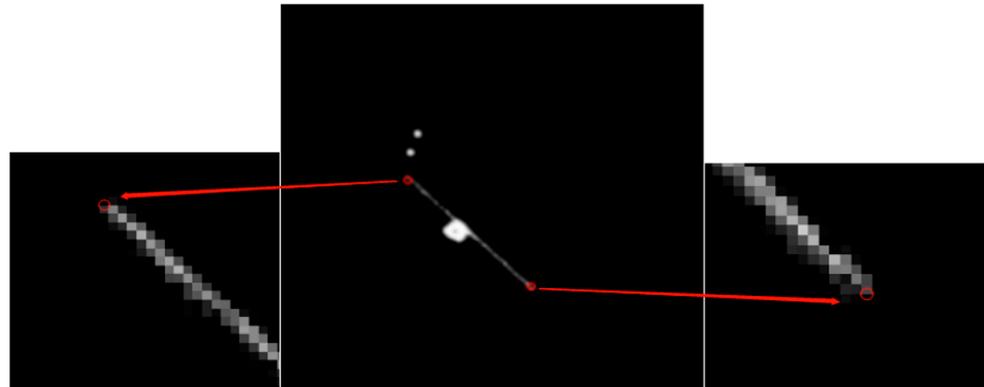


Figure 2. Schematic diagram of endpoint localization of spatial target images.

3. Super-Resolution Network of Space Target Image

3.1. Network Structure

In this paper, the design of the super-resolution network is constructed based on the U-Net network [19], and the network structure is shown in Figure 3, which mainly contains two parts: the primal regression network and the dual regression network. The part indicated by the black arrow in the model is the primal network of the model, while the part indicated by the red arrow corresponds to the dual regression network.

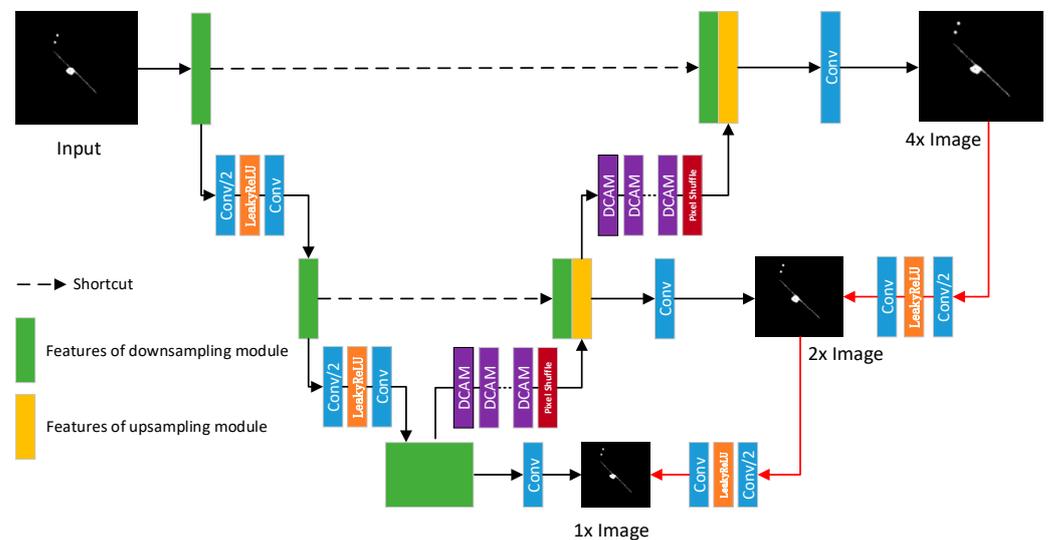


Figure 3. Diagram of overall network structure.

The primal regression network consists mainly of downsampling blocks and upsampling blocks. The downsampling block uses a series of convolutional layers with a step size of 2, a *LeakyReLU* activation function, and a convolutional layer with a step size of 1. This downsampling block is capable of extracting more complex and detailed information using pixel-level modeling capabilities. The upsampling base block consists of B deformable convolutional attention modules and sub-pixel convolutional layers, giving the network a more powerful feature representation and relevant feature learning capability, allowing the network to focus more on relevant features with differentiation and extract richer feature-vector-related information during training. Finally, the same structure as the downsampling module of the primal regression network is used to form a dual regression

network, which forms a closed loop with the primal regression network, and the two feed the generated information to each other for training.

3.2. Dual Regression Network

At this stage, most image SR reconstruction networks contain only the primal regression task, that is, the mapping relationship from LR to HR, but image super-resolution is a typical ill-posed problem where the mapping relationship between LR images and HR images has an uncertain nature; that is, there exist an infinite number of HR images that can be downsampled to obtain the same LR images, which makes the mapping space from LR to HR too large and the model appear to have the problem of self-adaptation. The dual regression network in this paper can solve this problem well, and the network contains both LR-to-HR mapping relations and HR-to-LR mapping relations.

Dual learning [20] was originally proposed in machine translation tasks to solve the problem of insufficient data for model training and has since been widely used in supervised learning tasks such as machine translation, sentiment analysis, image processing, and problem generation. Dual learning is achieved by giving a primal task model x , and a dual task model y provides feedback to model x ; similarly, given a dual task model y , its primal task model x can also provide feedback to the model y . The network framework is shown in Figure 4.

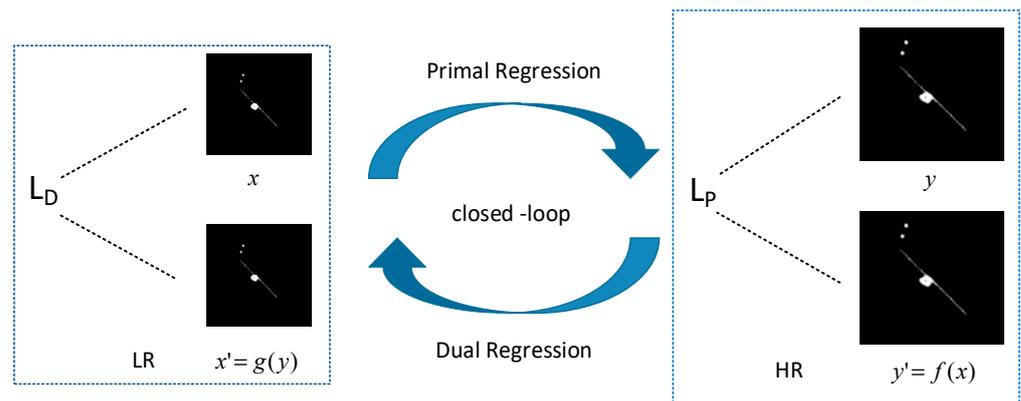


Figure 4. Schematic diagram of dual regression network structure.

3.3. Deformable Convolutional Attention Module Design

3.3.1. Attentional Mechanism

Inspired by the Convolutional Block Attention Module (CBAM) algorithm [21], this paper introduces channel attention [22] and spatial attention [23], where the channel attention module makes the model focus on meaningful information relevant to a specific task while suppressing interference from irrelevant information. The spatial attention module, on the other hand, focuses on which information at which locations is relevant to a particular task, while ignoring information at irrelevant locations. The combination of these two modules gives the model the ability to focus on both what and where.

The channel attention structure is shown in Figure 5, where the image $F(H \times W \times C)$ is channel dimension-weighted and first undergoes the maximum pooling and average pooling operations based on H and W to obtain two $1 \times 1 \times C$ outputs, which are then fed into the shared connection layer MLP and processed separately to obtain two feature maps according to the $ReLU$ activation function, followed by the sum operation and multiplication operation with *element-wise* to obtain the channel attention feature Mc . The whole process of channel attention operation can be expressed as

$$Mc(F) = \sigma(MLP(AvgPool(F)) + MLP(MaxPool(F))), \tag{6}$$

where σ is the *Sigmoid* activation function, *MLP* is the multilayer perceptron, *AvgPool* is the average pooling, and *MaxPool* is the maximum pooling.

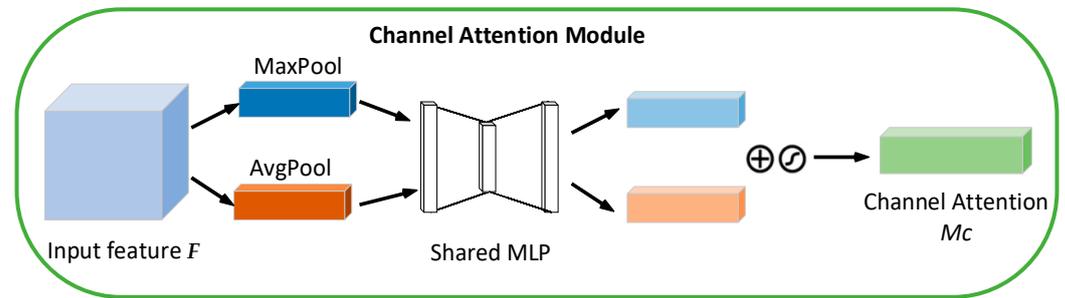


Figure 5. Structure diagram of channel attention module.

The structure of the spatial attention mechanism is shown in Figure 6. F' is the input image of the spatial attention module, and the size remains $H \times W \times C$. First, the two $H \times W \times 1$ outputs were obtained by carrying out two pooling operations based on the channels, and then downscaling them into one $H \times W \times 1$ output, and after *Sigmoid* activation processing, the output feature M_s and the input feature map F' were multiplied to obtain the final weighted features for learning. The whole operation process of spatial attention can be expressed as follows:

$$M_s(F') = \sigma\left(f^{7 \times 7}([AvgPool(F'); MaxPool(F')])\right), \tag{7}$$

where σ is the *Sigmoid* operation, $f^{7 \times 7}$ is the convolution operation, and the size of the convolution kernel is 7×7 .

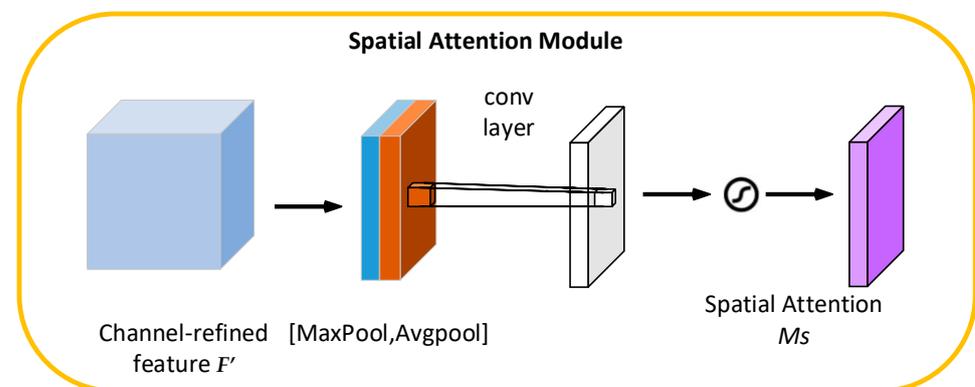


Figure 6. Structure diagram of the spatial attention module.

3.3.2. Deformable Convolution

For the standard convolution process, the output feature map y at each position $y(P_0)$ is calculated as

$$y(P_0) = \sum_{P_n \in R} \omega(P_n)x(P_0 + P_n), \tag{8}$$

where P_n are all sampled positions in R and P_0 is each position in the input feature map. The deformable convolution [24] process is given by

$$y(P_0) = \sum_{P_n \in R} \omega(P_n)x(P_0 + P_n + \Delta P_n), \tag{9}$$

where ΔP_n is the sampling point offset.

Since the position after adding the offset is non-integer and does not correspond to the actual pixel point on the feature map, it is necessary to use interpolation to obtain the pixel value after the offset, which can usually be performed using bilinear interpolation.

As can be seen, deformable convolution is the addition of a sampling point offset ΔP_n to the traditional convolution operation to adjust the sampling position of key elements. Deformable convolution only adds a small number of parameters and calculations to the neural network mode, but greatly improves the extraction of high-frequency features.

3.3.3. DCAM Module

As the number of network layers deepens, the network can perform more complex feature extraction, but at the same time, the model is prone to overfitting and gradient disappearance and explosion, etc. The introduction of the residual learning [25] network structure to learn the feature mapping was simpler than that of the deep network model, which greatly improved the performance of the deep network model. In this paper, the Deformable Convolutional Attention Module (DCAM) was constructed based on the residual module used in the RCAN [26] model, which fuses the channel attention mechanism and the spatial attention mechanism and uses deformable convolution instead of normal convolution to effectively extract the high-frequency information in the feature images. The structure diagram of the improved Deformable Convolutional Attention Module (DCAM) is shown in Figure 7.

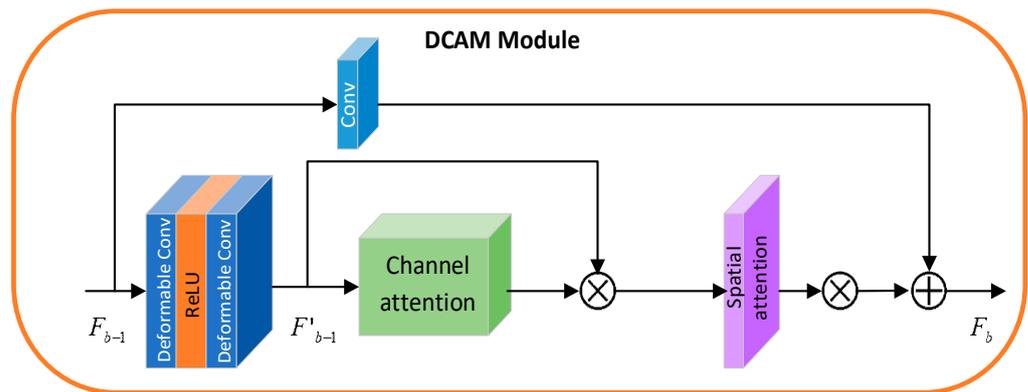


Figure 7. Structure diagram of DCAM module.

The output of the b th DCAM is shown in Equation (10):

$$F_b = H_b(F_{b-1}), \tag{10}$$

where F_{b-1} and F_b are the input and output of the b th DCAM, respectively, and H_b denotes the b th DCAM function, with the following operational details:

First, F_{b-1} is sequentially passed through the deformable convolutional layer, $ReLU$ activation layer, and deformable convolutional layer for feature extraction to obtain F'_{b-1} , which is calculated as shown in Equation (11):

$$F'_{b-1} = W_R^2 \delta \left(W_R^1 (F_{b-1}) \right), \tag{11}$$

where W_R^2 and W_R^1 represent the two deformable convolutions, respectively, and δ represents the $ReLU$ activation layer.

Next, F'_{b-1} is sequentially passed through the channel attention module and the spatial attention module to obtain the generated feature maps. Finally, F_b is obtained by summing it with the original input, which is calculated as shown in Equation (12):

$$F_b = U_{CA}(F'_{b-1}) \otimes U_{SA}(F'_{b-1}) + F_{b-1}, \tag{12}$$

where U_{CA} and U_{SA} represent the channel attention module and the spatial attention module, respectively, and \otimes is noted as *element – wise* multiplication.

3.4. Loss Function

The loss function has two main components: the loss of the primal regression network and the loss of the dual regression network. Given a set of N pairs of samples $S_p = \{(xi, yi)\}_{i=1}^N$, (xi, yi) denotes the i th pair of LR and HR images in this set of paired data. The training loss L is shown in Equation (13):

$$L = \sum_{i=1}^N L_p\{P(xi), yi\} + \lambda L_D\{D(P(xi)), xi\}, \quad (13)$$

where $P(xi)$ is the SR image predicted by the initial model and $D(P(xi))$ is the LR image obtained by downsampling from the dual model. L_p and L_D are the primal reconstruction network loss and the dual regression network loss, respectively. λ is the weight to control the percentage of dual regression loss.

4. Experimental Analysis

4.1. Experimental Dataset

In the experiment, the images taken by a telescope from different angles and different periods in “stellar gazing mode” were used as datasets. The images taken by the telescope in “stellar gazing mode” contained space targets and a certain number of stars; the stars appeared as dots, and the space target was in the shape of a dashed line, and the image format of the dataset was tif, with a total of 2000 images.

The dataset was randomly divided into a training set and a test set; 80% was selected as the training set and 20% as the test set. Because we cannot know the real position of the space target when calculating positioning coordinates, the test set and the simulated space target image were superimposed as true values, as shown in Figure 8. In addition, the training data were randomly rotated by 90° , 180° , and 270° , as well as panned and flipped using data enhancement techniques [27].



Figure 8. Example image of space target image test dataset.

The experiments were conducted with SR reconstruction at $2\times$ and $4\times$. During training and testing, the original image was first downscaled and then fed into the network for amplification and reconstruction, and the downscaling method used a bicubic interpolation method to reduce the original image resolution to $1/2$ and $1/4$ of the original, and the experiments for the two resolution amplifications were conducted independently.

4.2. Evaluation Indicators

Peak signal-to-noise ratio (PSNR) [28,29] and structural similarity index (SSIM) [30] are commonly used as metrics for the objective evaluation of image super-resolution. PSNR

indicates the ratio of the maximum power of the signal to the noise power that can affect its representation accuracy. A larger PSNR value indicates that the pixel difference between the generated image and the original image is smaller and the reconstruction effect is better. The SSIM reflects the structural similarity between the two images, and the closer the value is to 1, the better the reconstruction effect.

$$PSNR = 10 \cdot \log_{10} \left(\frac{M_{\max}^2}{E_{MS}} \right), \quad (14)$$

$$SSIM = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)}, \quad (15)$$

where M_{\max} denotes the maximum signal value present in HR, E_{MS} denotes the mean square error between SR images and HR images, μ_x and μ_y denote the mean of x and y images, σ_x^2 and σ_y^2 denote the variance of x and y , and σ_{xy} denotes the covariance of x and y ; C_1 and C_2 are constant values used to maintain stability.

In addition, in the algorithm of the super-resolution reconstruction of the space target, the evaluation index also includes the accurate calculation of the endpoint localization of the reconstructed space target, which is also the premise of trajectory calculation of the space target.

4.3. Model Details

The experiments in this paper are based on the PyTorch framework, using an NVIDIA GeForce 3090 graphics card for network training with 24 GB of video memory. The optimizer used the Adam optimizer, where $\beta_1 = 0.1$, $\beta_2 = 0.99$, and the batch size was set to 32. The learning rate was initialized to 10^{-4} and reduced to 10^{-7} by the cosine annealing algorithm. B was the number of DCAMs and F the number of underlying feature channels, set $B = 30$, $F = 16$, and the weight coefficient of the dual regression loss function λ was set to 0.1.

4.4. Ablation Experiments

To study the effectiveness of the attention module and the deformable convolution module, different attention modules and the deformable convolution module were combined to carry out comparative experiments. The experimental data are shown in Table 1, where parameters refer to how many parameters the model contains. PSNR and SSIM are mean values calculated from the dataset. The baseline does not integrate any modules.

Table 1. Comparative experimental data of attention module and deformable convolution module.

Comparison Item	Baseline	1	2	3	4	5	6	7
Channel Attention (CA)	×	√	×	×	×	√	×	×
Spatial Attention (SA)	×	×	√	×	×	×	√	×
CA + SA	×	×	×	√	×	×	×	√
Deformable Convolution Module	×	×	×	×	√	√	√	√
Parameters (M)	0.61	4.80	2.28	5.06	2.38	6.62	4.48	6.94
PSNR (dB)	54.86	54.99	54.96	55.74	54.82	55.70	55.68	55.96
SSIM	0.9918	0.9924	0.9920	0.9930	0.9925	0.9930	0.9933	0.9936

4.5. Experimental Results and Analysis

Bicubic [14], SRCNN [17], RCAN [26], and DRN [31] were selected for comparison with the algorithm in this paper under the same experimental settings with the magnifications of $\times 2$ and $\times 4$, respectively, and the experimental results are shown in Table 2.

Table 2. Comparison of PSNR and SSIM values of different algorithms.

Method	Scale	PSNR (dB)	SSIM
Bicubic	×2	47.84	0.9904
SRCNN		51.42	0.9875
RCAN		52.98	0.9917
DRN		54.99	0.9933
Ours		55.96	0.9936
Bicubic	×4	47.83	0.9903
SRCNN		50.23	0.9855
RCAN		51.59	0.9905
DRN		54.56	0.9906
Ours		54.99	0.9933

As can be seen from Table 2, the objective evaluation indexes of the reconstructed images by the algorithm in this paper have obvious advantages at a two times scale magnification. At a four times scale magnification, although the performance of each algorithm decreases due to the increase in the scale factor, the objective results show that the performance of the algorithm in this paper is still better than other algorithms.

According to the identification and localization technology of the space target introduced in Section 2.2, the coordinates of the endpoints of the SR-reconstructed space target images were compared with the unprocessed space target images to calculate the average error of the localization of the two endpoints, and the experimental results are shown in Table 3.

Table 3. Average localization error of space target endpoints after reconstruction by different algorithms (pixels).

Scale	Endpoints	Coordinate	Bicubic	SRCNN	RCAN	DRN	Ours
×2	Starting point	x	0.4895	0.4392	0.4150	0.1114	0.1080
		y	0.3416	0.3178	0.3009	0.1100	0.0818
	Ending point	x	0.1952	0.1669	0.1544	0.1056	0.1013
		y	0.5620	0.3207	0.2653	0.1255	0.1180
×4	Starting point	x	0.5000	0.4500	0.4200	0.2222	0.1100
		y	0.3429	0.3257	0.3158	0.1644	0.1128
	Ending point	x	0.2000	0.1882	0.1569	0.1382	0.1024
		y	0.5644	0.3298	0.2666	0.2574	0.1280

From the calculation results, it can be seen that the endpoint localization of the reconstructed space target image had the smallest localization error of the reconstructed image by the algorithm in this paper at the two times scale magnification. At the four times scale magnification, although the localization error of each algorithm had increased due to the increase in the scale factor, the objective results showed that the localization accuracy of the algorithm in this paper was better than that of other algorithms.

5. Discussion

In this paper, we propose a super-resolution reconstruction network for space target images based on dual regression and the deformable convolutional attention mechanism. The experimental results show that the method outperforms the comparison algorithm in both the objective quality evaluation index and localization accuracy on the space target image dataset compared with the current mainstream image super-resolution algorithms. The precise positioning of endpoints of a space target can accurately describe the position of the target and calculate its angular velocity in the field of view, which is an important reference for the subsequent research into telescope attitude determination and target orbit estimation.

Only the super-resolution reconstruction of space target images up to four times has been studied so far, but higher image resolution is certainly more valuable for practical applications, so the next step will be to study higher magnification super-resolution reconstruction in combination with the a priori knowledge of space target images. In addition, only the field of space target images has been studied in this paper so far, but the DCAM module proposed in this paper can better extract the high-frequency information of images, which is also applicable to feature extraction of different types of images in other fields, and will be further extended in combination with images in other fields next.

6. Conclusions

In this paper, we propose an image super-resolution reconstruction network based on dual regression and the deformable convolutional attention mechanism for the super-resolution reconstruction of degraded, low-resolution space target images. The experimental results show that the method proposed in this paper performs well in the super-resolution reconstruction of space target images, achieves a clear reconstruction of space target images, reduces the artifacts of target images, enriches the image details, reduces the localization errors, improves the localization accuracy, and has great potential in the field of super-resolution of space target images.

Author Contributions: Conceptualization, Y.S., C.J. and C.L.; methodology, Y.S. and Z.W.; software, Y.S. and C.J.; data curation, C.J. and C.L.; writing—original draft preparation, Y.S. and Z.W.; writing—review and editing, Y.S., C.J., C.L., W.L. and Z.W.; investigation, C.L. and W.L.; supervision, C.J., W.L., C.L. and Z.W.; validation, Y.S. and C.J.; funding acquisition, C.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by a special project for the high-technology industrialization of science and technology cooperation between Jilin Province and the Chinese Academy of Sciences (grant number E20833U9E0), regarding short-wave infrared sensor depth cooling vacuum encapsulation technology.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Meinel, A. Cost-Scaling Laws Applicable to Very Large Optical Telescopes. *J. Opt. Eng.* **1979**, *18*, 186645. [CrossRef]
2. van Belle, G.; Meinel, A.; Meinel, M. The Scaling Relationship between Telescope Cost and Aperture Size for Very Large Telescopes. SPIE. 2004, Volume 5489. Available online: <https://www.spiedigitallibrary.org/conference-proceedings-of-spie/5489/0000/The-scaling-relationship-between-telescope-cost-and-aperture-size-for/10.1117/12.552181.short?SSO=1> (accessed on 1 April 2023).
3. Harris, J.L. Diffraction and Resolving Power. *J. Opt. Soc. Am.* **1964**, *54*, 931–936. [CrossRef]
4. Tsai, R.Y.; Huang, T.S. *Multiframe Image Restoration and Registration*; JAI Press: Greenwich, CT, USA, 1984.
5. van Ouwerkerk, J.D. Image super-resolution survey. *Image Vis. Comput.* **2006**, *24*, 1039–1052. [CrossRef]
6. Greenspan, H. Super-Resolution in Medical Imaging. *Comput. J.* **2009**, *52*, 43–63. [CrossRef]
7. Isaac, J.S.; Kulkarni, R. Super resolution techniques for medical image processing. In Proceedings of the 2015 International Conference on Technologies for Sustainable Development (ICTSD), Mumbai, India, 4–6 February 2015; pp. 1–6.
8. Lin, F.; Fookes, C.; Chandran, V.; Sridharan, S. Super-Resolved Faces for Improved Face Recognition from Surveillance Video. In Proceedings of the International Conference on Biometrics, Seoul, Republic of Korea, 27–29 August 2007.
9. Yang, D.; Li, Z.; Xia, Y.; Chen, Z. Remote sensing image super-resolution: Challenges and approaches. In Proceedings of the 2015 IEEE International Conference on Digital Signal Processing (DSP), Singapore, 21–24 July 2015; pp. 196–200.
10. Dodgson, N.A. Quadratic interpolation for image resampling. *IEEE Trans. Image Process.* **1997**, *6*, 1322–1326. [CrossRef] [PubMed]
11. Hsieh, H.; Andrews, H. Cubic splines for image interpolation and digital filtering. *IEEE Trans. Acoust. Speech Signal Process.* **1978**, *26*, 508–517. [CrossRef]
12. Schultz, R.R.; Stevenson, R.L. A Bayesian approach to image expansion for improved definition. *IEEE Trans. Image Process.* **1994**, *3*, 233–242. [CrossRef] [PubMed]
13. Keys, R. Cubic convolution interpolation for digital image processing. *IEEE Trans. Acoust. Speech Signal Process.* **1981**, *29*, 1153–1160. [CrossRef]
14. Xin, L.; Orchard, M.T. New edge-directed interpolation. *IEEE Trans. Image Process.* **2001**, *10*, 1521–1527. [CrossRef] [PubMed]
15. Kim, S.P.; Bose, N.K.; Valenzuela, H.M. Recursive reconstruction of high resolution image from noisy undersampled multiframe. *IEEE Trans. Acoust. Speech Signal Process.* **1990**, *38*, 1013–1027. [CrossRef]

16. Freeman, W.T.; Jones, T.R.; Pasztor, E.C. Example-based super-resolution. *IEEE Comput. Graph. Appl.* **2002**, *22*, 56–65. [[CrossRef](#)]
17. Dong, C.; Loy, C.C.; He, K.; Tang, X. Image Super-Resolution Using Deep Convolutional Networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2016**, *38*, 295–307. [[CrossRef](#)] [[PubMed](#)]
18. Harris, C.G.; Stephens, M.J. A Combined Corner and Edge Detector. In Proceedings of the Alvey Vision Conference, Manchester, UK, 31 August–2 September 1988.
19. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015*; Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F., Eds.; Springer International Publishing: Cham, Switzerland, 2015; pp. 234–241.
20. Zhu, S.; Cao, R.; Yu, K. Dual Learning for Semi-Supervised Natural Language Understanding. *IEEE/ACM Trans. Audio Speech Lang. Process.* **2020**, *28*, 1936–1947. [[CrossRef](#)]
21. Woo, S.; Park, J.; Lee, J.-Y.; Kweon, I.S. CBAM: Convolutional Block Attention Module. In Proceedings of the Computer Vision—ECCV 2018, Munich, Germany, 8–14 September 2018; Ferrari, V., Hebert, M., Sminchisescu, C., Weiss, Y., Eds.; Springer International Publishing: Cham, Switzerland, 2018; pp. 3–19.
22. Hu, J.; Shen, L.; Sun, G. Squeeze-and-Excitation Networks. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 7132–7141.
23. Jaderberg, M.; Simonyan, K.; Zisserman, A.; Kavukcuoglu, K. Spatial transformer networks. In Proceedings of the 28th International Conference on Neural Information Processing Systems, Montreal, QC, Canada, 7–12 December 2015; MIT Press: Montreal, QC, Canada, 2015; Volume 2, pp. 2017–2025.
24. Dai, J.; Qi, H.; Xiong, Y.; Li, Y.; Zhang, G.; Hu, H.; Wei, Y. Deformable Convolutional Networks. In Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017; pp. 764–773.
25. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
26. Zhang, Y.; Li, K.; Li, K.; Wang, L.; Zhong, B.; Fu, Y. *Image Super-Resolution Using Very Deep Residual Channel Attention Networks*; Springer International Publishing: Cham, Switzerland, 2018; pp. 294–310.
27. Li, Z.; Yang, J.; Liu, Z.; Yang, X.; Jeon, G.; Wu, W. Feedback Network for Image Super-Resolution. In Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 15–20 June 2019; pp. 3862–3871.
28. Wang, Z.; Simoncelli, E.P.; Bovik, A.C. Multiscale structural similarity for image quality assessment. In Proceedings of the Thirty-Seventh Asilomar Conference on Signals, Systems & Computers, Pacific Grove, CA, USA, 9–12 November 2003; Volume 2, pp. 1398–1402.
29. Welstead, S.T. *Fractal and Wavelet Image Compression Techniques*; Society of Photo-Optical Instrumentation Engineers (SPIE): Bellingham, WA, USA, 1999.
30. Liu, Y.; Zhu, L.; Lim, K.; Li, Y.; Wang, F.; Lu, J. Review and Prospect of Image Super-Resolution Technology. *J. Front. Comput. Sci. Technol.* **2020**, *14*, 181–199.
31. Guo, Y.; Chen, J.; Wang, J.; Chen, Q.; Cao, J.; Deng, Z.; Xu, Y.; Tan, M. Closed-Loop Matters: Dual Regression Networks for Single Image Super-Resolution. In Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 13–19 June 2020; pp. 5406–5415.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.