A Method for Identifying the Number of Spiral Arms in Spiral Galaxies with an Integrated Attention Mechanism[*]

Du Ping¹   Dong Shuyu²   Zhang Jinqu^{2†}

2026-06-03

ChinaRxiv

Abstract

Accurately identifying the number of spiral arms in spiral galaxies using deep learning methods plays an important role in tracing galaxy evolutionary histories and understanding the evolutionary patterns of spiral structures across different cosmic epochs. However, current deep learning models do not fully consider the characteristics of spiral galaxy images, particularly the brightness distribution features of galactic centers and spiral arms. To address this issue, a brightness-based bidirectional fusion Manhattan attention mechanism is proposed and integrated with multiple network models such as EfficentNet to construct a spiral arm number identification method for spiral galaxies incorporating an attention mechanism. Taking spiral galaxy images selected from the Galaxy Zoo 2 dataset as examples, experimental comparative analysis was conducted. The results show that after adding the attention mechanism, the EfficientNet model achieves an overall accuracy of 95.52% on the test dataset, with particularly improved recognition performance for spiral arms numbering 3, 4, and above. This study demonstrates that a well-designed attention mechanism can mine the brightness distribution features of spiral galaxies and improve the effectiveness of spiral arm number identification.

Full Text

doi: 10.15940/j.cnki.0001-5245.2026.03.004

A Method for Identifying the Number of Spiral Arms in Spiral Galaxies with an Integrated Attention Mechanism*

Du Ping1 Dong Shuyu2 Zhang Jinqu2†

(1. School of Architectural Information, Guangdong Construction Polytechnic, Qingyuan 511500)
(2. School of Computer Science, South China Normal University, Guangzhou 510631)

Abstract The use of deep learning methods to accurately identify the number of spiral arms in spiral galaxies is of great importance for tracing the evolutionary trajectories of galaxies and understanding the evolutionary regularities of spiral structures across different cosmic epochs. However, current deep learning models do not sufficiently take into account the characteristics of spiral-galaxy images, especially the distribution features of the brightness of the galactic center and spiral arms. To address this issue, a brightness-based bidirectional fusion Manhattan attention mechanism is proposed. By combining this attention mechanism with multiple network models such as EfficientNet, a method for identifying the number of spiral arms in spiral galaxies with an integrated attention mechanism is constructed. Using spiral-galaxy images selected from the Galaxy Zoo 2 dataset as examples, comparative experiments are carried out. The results show that, after the attention mechanism is added, the EfficientNet model achieves an overall accuracy of 95.52% on the test dataset, with particularly better recognition performance for galaxies with 3, 4, or more spiral arms. The study shows that a well-designed attention mechanism can mine the brightness-distribution features of spiral galaxies and improve the effectiveness of spiral-arm number recognition.

Keywords galaxies: spiral galaxies; techniques: image processing; methods: data analysis; methods: classification

Chinese Library Classification Number: P157; **Document Code:** A

1 Introduction

Deep learning, as an important branch of machine learning, has been widely applied in fields such as galaxy morphology classification, stellar spectral-parameter estimation, and pulsar signal processing[1-5]. With the continued development of research, the application of deep learning in astronomical data processing has also been continuously expanding and deepening. In recent years, how to accurately identify the number of spiral arms in spiral galaxies through deep learning methods has become an important direction[6-7].

Judging from current methods for identifying the number of spiral arms, researchers mainly achieve the recognition of different numbers of spiral arms through different deep learning networks. For example, Lee et al. used different transfer-learning schemes to construct different EfficientNet models for classifying the number of spiral arms, ultimately obtaining the best F1-score on the EfficientNetV2-M architecture[8]. Davis et al. proposed the machine-vision algorithm SpArcFiRe (the SPiral ARC FInder and REporter), which analyzes galaxy images to realize automatic identification and segmentation of spiral-arm segments, and further calculates parameters such as the pitch angle and length of spiral arms[9]. Hart et al. selected a stellar-mass complete sample of 6222 spiral galaxies from the Sloan Digital Sky Survey (SDSS) Galaxy Zoo 2 dataset, and used the SpArcFiRe algorithm to identify spiral-arm features and measure their related geometric shapes[10]. In addition, Deutsch et al. trained six different deep learning models based on the U-Net architecture on the Galaxy Zoo dataset to detect the spiral arms of spiral galaxies[11]. Through analysis of current applications of deep learning to spiral-arm number recognition, it can be seen that deep learning is an effective method for spiral-arm recognition. At the same time, it can also be seen that the current level of application of deep learning technology remains limited to the selection and application of existing network architectures, and the generalization ability of the models is limited. In particular, regarding how to optimize networks based on the characteristics of spiral-galaxy images so as to further improve recognition performance, research

Received original manuscript on 2025-06-11; revised manuscript received on 2025-08-25.

* Supported by the China Manned Space Engineering Survey Space Telescope Special Project (CMS-CSST-2025-A02) and the Guangdong Provincial Department of Education Teaching Reform Project (2023JG067).
† zjq@scnu.edu.cn

is still not sufficiently in-depth.

Spiral galaxy images have distinct spiral-arm features, and the brightness distribution in such images therefore follows special distribution patterns [12]. How to use the brightness-distribution characteristics of spiral galaxy images to optimize the performance of network models in recognizing the number of spiral arms is of great importance. The key to improving model performance lies in paying particular attention, during computation, to the spatial distribution of spiral arms. The attention mechanism, as a mechanism that imitates the "focused attention" of human visual and cognitive systems, is implemented by computing the weights of different parts of the input data and dynamically adjusting the allocation of computational resources in the model, thereby improving the efficiency and performance of the model [13]. For example, the Transformer model adopts the self-attention mechanism, enabling the model to dynamically attend to words at different positions in an input sentence and thus capture semantic relationships more accurately [14]. Models such as ViT (Vision Transformer) introduce attention mechanisms, allowing the network to focus on key regions of an image [15]. Today, attention mechanisms have become a key technology in deep-learning architectures and have been widely applied in many fields, including natural language processing, computer vision, and speech recognition.

Existing attention mechanisms are mainly divided into categories such as channel attention, spatial attention, and self-attention. In application design, they mostly employ global-attention expansion or attention-judgment strategies based on the image center. For spiral galaxy images, there is a special spatial positional relationship between the locations of spiral arms and the galaxy center. Meanwhile, multiple bright regions may exist in spiral galaxy images, and the brightest region is not necessarily located at the image center. Different attention mechanisms will have an important impact on spiral-arm recognition. Therefore, this paper attempts to construct an attention mechanism adapted to spiral galaxy images and to build a spiral-arm number recognition model incorporating the attention mechanism.

2 Method

2.1 Basic idea

Since a spiral galaxy image is a type of two-dimensional data, and the spiral arms exhibit varying brightness changes outward from the spiral center, the design of the attention mechanism requires an attenuation mechanism that can effectively capture this spatial distribution. The Manhattan Self-Attention (MaSA) mechanism is precisely an attention mechanism that transforms unidirectional one-dimensional temporal attenuation into bidirectional two-dimensional spatial attenuation [16]. Therefore, in designing the attention mechanism, this paper takes the MaSA attention mechanism as the basis and, in combination with the characteristics of spiral galaxy images, designs a brightness-based Manhattan Distance Spatial Attention (BMSA) mechanism.

2.2 Design of a bidirectional fusion attention mechanism based on Manhattan distance

MaSA adopts spatial attenuation related to Manhattan Distance, that is, the distance between two positions is the sum of the horizontal and vertical distances between two points on a two-dimensional plane. This attenuation mechanism allows spatially nearby regions to retain more information, while farther regions undergo greater attenuation. In the task of classifying spiral arms in spiral galaxies, Manhattan distance can effectively measure the relative position of a pixel with respect to the image center and can effectively simulate the geometric distribution of spiral arms. Since spiral arms usually extend gradually from the center of the image to the periphery, using Manhattan distance to capture the spatial relationship between pixels and the center enables the model to adaptively weight spiral arms in space. At the same time, to reflect the characteristic that the brightness distribution of spiral arms decays from the central line of the spiral arm toward both sides, this paper simultaneously sets a forward attenuation parameter and a backward attenuation parameter in the attention-mask design, introduces a fusion coefficient α for them, and constrains this coefficient α by setting a sigmoid function so that it always remains between 0 and 1, thereby ensuring that the model can adaptively adjust it during training. The principle of the bidirectional fusion attention mechanism is shown in Figure 1.

Figure 1 Schematic diagram of the principle of the bidirectional fusion Manhattan attention mechanism

Fig. 1 Principle diagram of bidirectional fusion with Manhattan attention mechanism

In Figure 1, λ_{forward} denotes the forward attenuation parameter; $\lambda_{\text{backward}}$ denotes the backward attenuation parameter; and $\sigma(\alpha)$ denotes the output after the sigmoid activation function. It can be seen from the figure that, after bidirectional fusion, the attention is mainly concentrated in the middle region of the image. This strengthens the model' s attention to the positions of spiral arms and increases the weights of the relevant regions, thereby helping the model better recognize detailed information about the spiral arms. The specific formulas for forward attenuation attention (Forward Attention), backward attenuation attention (Backward Attention), and the attention mask (Attention Mask) are as follows:

$$d_{\text{manhattan}} = \left|y_{\text{coord}} - \text{center}_y\right| + \left|x_{\text{coord}} - \text{center}_x\right|, \tag{1}$$

$$\text{ForwardAttention} = \exp\left(-\lambda_{\text{forward}} \cdot d_{\text{manhattan}}\right), \tag{2}$$

$$\text{BackwardAttention} = 1 - \exp\left(-\lambda_{\text{backward}} \cdot d_{\text{manhattan}}\right), \tag{3}$$

$$\begin{aligned} \text{AttentionMask} = {} & \sigma(\alpha) \cdot \text{ForwardAttention} \\ & + [1 - \sigma(\alpha)] \cdot \text{BackwardAttention}, \end{aligned} \tag{4}$$

where $d_{\text{manhattan}}$ denotes the Manhattan distance calculated for a certain point on the image; x_{coord} and y_{coord} denote the x- and y-coordinates of a pixel point in the image, respectively; and center_x and center_y denote the x- and y-values of the image center, respectively. By performing weighted fusion of the forward decaying attention ForwardAttention and the backward decaying attention BackwardAttention, the final attention mask AttentionMask is obtained. This mask computed by the attention mechanism is incorporated into the model, enabling the model to focus on capturing the shape and distribution features of spiral arms, thereby improving the accuracy of spiral-arm localization and classification.

2.3 Design of a Brightness-Based Bidirectional Fusion Manhattan Attention Mechanism

Although the bidirectional fusion attention mechanism improves the capture of the spatial positions of spiral arms, it mainly uses the image center as the reference to calculate the attention at different positions, whereas the image center is not necessarily the center of the spiral galaxy. To effectively identify the main parts of a spiral galaxy—especially the galaxy center and spiral-arm positions—this study further calculates the center coordinates of the spiral galaxy through brightness-threshold filtering, thereby optimizing the application of the attention mechanism.

First, the brightness map of the spiral-galaxy image is calculated by averaging the RGB channels of the image to obtain the brightness map. Because the brightness difference between the galaxy component and the sky background in galaxy images is usually large, the mean brightness is used as the threshold to binarize the bright regions in the image: regions with brightness greater than the mean are set to 1, and otherwise to 0. Through binarization, regions with higher brightness in the image are highlighted. The specific calculation formulas are as follows:

$$I_{\text{lum}}^{(b)}(x, y) = \frac{1}{C} \sum_{c=1}^{C} X(b, c, x, y), \ \forall b \in \{1, \cdots, T\}, \tag{5}$$

$$\mu_{\text{lum}}^{(b)} = \frac{1}{HW} \sum_{x=1}^{H} \sum_{y=1}^{W} I_{\text{lum}}^{(b)}(x, y), \tag{6}$$

$$B^{(b)}(x, y) = \begin{cases} 1 & \text{if } I_{\text{lum}}^{(b)}(x, y) > \mu_{\text{lum}}^{(b)}, \\ 0 & \text{otherwise.} \end{cases} \tag{7}$$

where T denotes the total number of images; C denotes the number of channels in a single image; $X(b, c, x, y)$ denotes the pixel value of the c-th channel of the b-th image at position (x, y); H denotes the image height, and W denotes the image width. In the calculation, the mean pixel value over all channels at each position (x, y) is first computed to obtain the single-channel brightness map $I_{\text{lum}}^{(b)}$. Then, the mean brightness of all pixels in image b is calculated to obtain the global brightness threshold $\mu_{\text{lum}}^{(b)}$. Finally, pixels whose brightness value $I_{\text{lum}}^{(b)}(x, y)$ is greater than the global brightness threshold are marked as 1, indicating high-brightness regions; otherwise, they are marked as 0, indicating background regions. The binary result $B^{(b)}$ of the image is thus obtained.

After binarizing the brightness map, connected-component analysis is performed on the image to identify the largest and brightest connected region in the image, which is judged to be the main region of the spiral galaxy. The center position of this region is calculated and passed to the BMSA attention mechanism, replacing the image-center coordinates used in the original attention mechanism. The specific process is shown in Fig. 2.

Original Image → Brightness Map → Binary Image → Marked Image

Fig. 2 Diagram of brightness center calculation

Compared with the traditional attention mechanism with a fixed center position, BMSA can dynamically adjust the center position, making the attention mechanism focus more on the most representative bright region in the image, thereby improving the model' s performance in spiral-galaxy recognition.

2.4 Construction of an EfficientNet Model Integrating the Attention Mechanism

To verify the effect of the attention mechanism designed in this paper, we

Using the EfficientNet model as an example, we focus on the design and application of attention integration. EfficientNet is an efficient convolutional neural network architecture proposed by Google Research in 2019. By jointly scaling the depth, width, and resolution of the network, it achieves optimal allocation of computational resources, thereby improving model performance while reducing computational cost. In 2022, Wu et al. designed the Efficient-G series networks for galaxy morphology classification based on EfficientNet, and performed morphological classification on galaxy images from Galaxy Zoo, achieving a classification accuracy close to 95% [17]. In this paper, a brightness-based bidirectional-fusion Manhattan attention mechanism is integrated into the Efficient-G3 network. The network structure and parameters of Efficient-G3 are shown in Table 1.

Table 1 EfficientNet-G3 Structure
Table 1 Structure of EfficientNet-G3

Stage	Operator	Resolution	#Channels	#Layers
1	Conv5 $\times$ 5	96×96	56	1
2	iRBlockConv1,k3 $\times$ 3	48×48	32	3
3	iRBlockConv6,k3 $\times$ 3	24×24	40	5
4	iRBlockConv6,k5 $\times$ 5	12×12	72	5
5	iRBlockConv6,k3 $\times$ 3	6×6	144	7
6	Conv1 $\times$ 1 &Pooling & FC	6×6	288	1

In the structure of EfficientNet-G3, Conv5 $\times$ 5 denotes a 5×5 convolution operation performed in that layer; iRBlock (Inverted Residual Block) is an inverted residual structure module. iRBlockConv6, k3 $\times$ 3 indicates that the number of convolution kernels in the first convolutional processing layer of the iRBlock module is 6, and that the convolution kernel size is 3×3; FC (Fully Connected Layer) denotes a fully connected layer. From its structure, it can be seen that the key component used by this network is the inverted residual block iRBlock. This module first expands the dimensionality through an Expand_conv 1×1 convolution operation, then applies depthwise separable convolution Depthwise_conv and the Squeeze-Excitation operation to each input channel, and finally performs Project_conv convolution to restore the number of channels and prepare for skip connection. We integrate the brightness bidirectional-fusion Manhattan attention mechanism before the Squeeze-Excitation operation. The improved inverted residual block structure is shown in Fig. 3.

Fig. 3 diagram labels: Inverted Residual Block; Expand_conv 1×1; Depthwise_conv $3 \times 3/5 \times 5$; Brightness Manhattan Spatial Attention; Squeeze Excitation 1×1; Project_conv 1×1.

Figure 3 iRBlock module structure after integrating the brightness bidirectional-fusion Manhattan attention mechanism
Fig. 3 iRBlock module structure after integrated brightness and bidirectional fusion with Manhattan attention mechanism

3 Experiments and Results

3.1 Experimental Data and Parameter Settings

The experimental data in this paper are mainly obtained by processing and organizing the Galaxy Zoo 2 [18] dataset. The Galaxy Zoo catalog decision tree provides three methods for describing the voting results of each task, namely vote fraction, weighted vote fraction, and debiased vote fraction. This paper follows the clean-sample selection rules of the Galaxy Zoo project, using the weighted vote fraction and the debiased vote fraction to screen the images. First

According to the weighted voting score (ρ), clean samples were screened, and spiral-galaxy images whose basic vote count (N) for the target task exceeded a certain threshold were selected. The specific selection rules are shown in Table 2. For the selected spiral-galaxy images, the number of spiral arms was divided into six classes according to the voting proportion for the number of spiral arms. Classes 1–4 are galaxy images with the corresponding number of spiral arms; Class 5 consists of resolvable galaxies with more than four spiral arms; and Class 6 consists of galaxy images for which the number of spiral arms cannot be accurately determined.

Table 2. Selection rules for clean samples of spiral galaxy images in Galaxy Zoo 2

Galaxy Type	Question	Voting points	Total
Spiral Galaxy	T01	$f_{\text{feature/disk}} > 0.430$	13019
Spiral Galaxy	T02	$f_{\text{edge-on,no}} > 0.715$	13019
Spiral Galaxy	T03	$f_{\text{spiral,yes}} > 0.715$	13019
Spiral Galaxy	T04	$N_{\text{spiral,yes}} > 20$	13019

According to the rules in Table 2, a total of 13,019 clean samples were selected. The number of spiral arms was then assigned using the classification condition that the voting proportion for the number of spiral arms was greater than 0.8. Among them, the largest category comprised galaxy images with two spiral arms, containing 10,910 images, whereas the smallest category comprised galaxies with four spiral arms, with only 77 images. There were 79 galaxies with more than four spiral arms, and 1,153 galaxies for which the number of spiral arms could not be counted. In addition, there were 206 and 594 galaxies with one and three spiral arms, respectively.

During the experiments, the dataset was divided into a training set and a test set at a ratio of 8:2. To ensure that both the training and test sets contained images with each number of spiral arms, the spiral-galaxy images with each arm count were kept in the same proportional distribution when the dataset was filtered. This study used Python 3.8 and called libraries such as Pandas, Torch, and

Scikit-learn. Computation was performed using an NVIDIA GeForce RTX 3080 Ti Laptop GPU. The training batch size was set to 32, the Adam optimizer and cross-entropy loss function were used, the learning rate was set to 0.0001, and experiments were conducted for 500 iterations. The specific experimental parameters are shown in Table 3. The complete experimental code is available in the git repository: https://gitee.com/zhangjin-district/EfficientNet_BMSA.

3.2 Results of Adding the BMSA Attention Mechanism for Identifying the Number of Spiral Arms in Spiral Galaxies

To verify the effect of the BMSA attention mechanism, this paper takes the EfficientNet-G3 model as an example and calculates the results before and after using the BMSA attention mechanism. The results are shown in Fig. 4. It can be seen from the figure that, after integrating the BMSA attention mechanism, all performance indicators of the EfficientNet-G3 model improved, indicating that the BMSA attention mechanism, to a certain extent, focuses on the brightness-distribution features of spiral galaxies.

Table 3. Settings of experimental parameters

Parameter	Value
Image center cropping size range	[160, 240]
Color disturbance intensity	0.2
Flip probability	0.5
Batch size	32
Learning rate	0.0001
Attention mechanism attenuation coefficient	0.1
Channel scaling factor	2.6
Depth scaling factor	4.0
EfficientNet-Dropout	0.2

Fig. 4. Performance comparisons before and after integrating BMSA attention mechanism

Table 4 lists the recognition performance for different numbers of spiral arms. From the recognition results, it can be seen that, for spiral-galaxy images with three or more spiral arms, the classification accuracy increased after adding the BMSA attention mechanism. The recall rate for spiral-galaxy images with 1–4 spiral arms improved slightly. Only the F1 scores for spiral-galaxy images with one spiral arm and for spiral-galaxy images whose number of spiral arms could not be identified showed no change. It can be seen that, for most spiral-galaxy images, the classification performance after adding the BMSA attention mechanism is better, making it possible to identify the number of spiral arms in spiral galaxies more clearly. Judging from the experimental results, the BMSA attention the attention mechanism effectively helped the model perform the spiral-arm recognition task. In particular, for spiral galaxy images with easily confused and imbalanced numbers of spiral arms—namely 3, 4, and greater than 4—all metrics improved. Meanwhile, considering that the category with 4 spiral arms has the smallest proportion, the recognition performance improved significantly after the BMSA attention mechanism was added. Figure 5 shows two spiral galaxy images with 4 spiral arms. When the model was not equipped with the attention mechanism, it recognized the image on the left as having 5 spiral arms and the image on the right as having 3 spiral arms, both of which were misclassifications. However, after the BMSA attention mechanism was added, both images were correctly recognized. It can thus be seen that the BMSA attention mechanism alleviates, to a certain extent, the recognition difficulties caused by the severe imbalance in the size of the spiral galaxy dataset.

Table 4 Analysis of the effectiveness of the EfficientNet model with and without the BMSA attention mechanism

Number of spiral arms	Precision	Precision	Recall	Recall	F1-Score	F1-Score
Number of spiral arms	EfficientNet-G3	+BMSA	EfficientNet-G3	+BMSA	EfficientNet-G3	+BMSA
1	86.11%	80.00%	77.50%	80.00%	81.58%	80.00%
2	97.72%	97.51%	97.63%	97.91%	97.68%	97.71%
3	88.99%	91.89%	79.51%	83.61%	83.98%	87.55%
4	64.29%	78.57%	52.94%	64.71%	58.06%	70.97%
>4	66.67%	73.86%	87.50%	87.50%	75.48%	80.00%
Uncountable	80.59%	81.98%	86.82%	86.82%	83.59%	82.35%

Figure 5 Case images with a spiral-arm number of 4

4 Discussion

4.1 Effect of the Attention-Mechanism Position on Model Performance

Deep-learning network architectures are generally composed of multiple layers, and applying an attention mechanism at different positions can also affect model performance. Therefore, the BMSA attention mechanism was inserted into different positions in the EfficientNet model to examine the influence of these positions on the model' s recognition of the number of spiral arms. In the experiments, positions before and after the Squeeze Excitation operation, before and after the iRBlock, and after the first convolutional layer were selected, respectively. The experimental results are shown in Table 5.

Table 5 The influence of the BMSA attention mechanism' s position on the recognition of the number of spiral arms

Location	Precision	Recall	F1-Score	Accuracy
Before Squeeze Excitation	**95.49%**	**95.52%**	**95.49%**	**95.52%**
Before iRBlock	95.19%	95.33%	95.20%	95.33%
After iRBlock	95.23%	95.33%	95.26%	95.33%
After Squeeze Excitation	87.62%	89.70%	88.15%	89.70%
After First Convolutional layer	93.68%	93.87%	93.72%	93.87%

The study found that adding the BMSA attention-mechanism module before the Squeeze Excitation module produced the best results, with all metrics achieving their optimal values. Placing BMSA before Squeeze Excitation allows spatial attention to first screen key regions, and then the Squeeze Excitation module optimizes channel weights; the two complement each other. Placing

The BMSA attention mechanism module performs worst when placed after the Squeeze Excitation module. The Squeeze Excitation module suppresses part of the spatial information through channel compression, weakening the brightness-difference features required by BMSA. In deep networks, the resolution of the feature map is relatively low and sensitivity to spatial information decreases. The performance before and after adding the BMSA attention mechanism module to the iRBlock module is relatively close. Adding it after the first convolutional layer yields relatively poor results, but it is still better than adding the BMSA attention mechanism after the Squeeze Excitation module. Overall, adding the BMSA attention mechanism before the Squeeze Excitation module gives the best effect; the model can use brightness differences to enhance attention to important regions, thereby improving overall performance.

4.2 Influence of Attention Spatial Attenuation Parameters on Spiral-Arm Number Identification

The forward attenuation parameter and reverse attenuation parameter affect the attention distribution of BMSA. Forward attenuation can strengthen the influence of high-brightness regions, whereas reverse attenuation can reduce the influence of low-brightness regions. To explore the effect of combining the two, this paper sets different attenuation parameter values such as 0.1, 0.3, and 0.5 for the EfficientNet model fused with BMSA, and also tests the case in which the parameters are set as learnable parameters for comparison. The results are shown in Table 6.

Table 6 Analysis of the results for different attenuation parameters in identifying the number of spiral arms

Attenuation parameter	Precision	Recall	F1-Score	Accuracy
0.1	**95.49%**	**95.52%**	**95.49%**	**95.52%**
0.3	95.38%	95.48%	95.39%	95.48%
0.5	95.33%	95.40%	95.31%	95.40%
0.1-learn	95.43%	95.44%	95.38%	95.44%
0.3-learn	95.05%	95.17%	95.05%	95.17%

The results show that when the fixed parameter value is 0.1, the recognition effect is the best, with an accuracy of 95.52%, an F1 score of 95.49%, a precision of 95.49%, and a recall of 95.52%, leading across all four indicators. Through ablation experiments, it can be seen that when the parameter values are set to 0.3 and 0.5, all indicators obtained in the experiment show a stepwise downward trend. To enhance the generalization ability of the model, the attenuation parameters are set to be determined through adaptive learning. Here, 0.1-learn and 0.3-learn indicate that learning starts from initial values of 0.1 and 0.3, respectively. After 500 iterations of learning, experiments are conducted. The results show that the parameter group initially set to 0.1 still maintains a high level, but is slightly lower than the group with the fixed parameter of 0.1.

4.3 Effect Analysis of the Attention Mechanism on Other Deep Learning Networks

To verify the effect of the attention mechanism proposed in this paper on different deep learning network structures, three different neural network structures—ResNet34, DenseNet[19], and RMT[16] (Retentive Networks Meet Vision Transformers)—are further compared after adding the BMSA attention mechanism. In the experiments, the hyperparameters of each network model adopt settings consistent with those of the preceding EfficientNet experiment.

It can be seen from Table 7 that after adding the BMSA attention mechanism, the three different network models all improve in terms of precision, recall, F1 value, and overall Accuracy. This indicates that the bidirectional fusion attention guided by brightness enables the model to capture galaxy structures more accurately and improves the model' s recognition performance. However, it can also be found that the attention mechanism has certain differences in effect across different network structures. Taking the Accuracy indicator as an example, DenseNet shows the greatest improvement. Even so, judging from the final results of several network models, the EfficientNet-G3 model integrated with the BMSA attention mechanism still achieves the relatively best results. The fundamental reason why the attention mechanism has different improvement effects on different network architectures lies in the considerable differences in the inherent structural characteristics and information-flow patterns of these networks.

Table 7 The effect of BMSA attention mechanism on different network structures

Model	Precision	Recall	F1-Score	Accuracy
ResNet34	94.37%	94.41%	94.27%	94.41%
ResNet34 + BMSA	94.66%	94.71%	94.61%	94.71%
DenseNet	93.68%	93.87%	93.72%	93.87%
DenseNet + BMSA	94.40%	94.48%	94.33%	94.48%
RMT	94.06%	94.27%	94.04%	94.27%
RMT + BMSA	94.42%	94.48%	94.40%	94.48%
EfficientNet-G3	95.28%	95.21%	95.21%	95.21%
EfficientNet-G3 + BMSA	**95.49%**	**95.52%**	**95.49%**	**95.52%**

4.4 Effect of Image Signal-to-Noise Ratio on Spiral-Arm Recognition Performance

The image signal-to-noise ratio (Signal-to-Noise Ratio, SNR) refers to the ratio of the image signal intensity to the noise intensity. A higher SNR value indicates better image quality, whereas a lower SNR value indicates that the image is subject to more noise interference. Because no clean images are available as references, this paper calculates the SNR by computing the mean and variance of the images. Fig. 6 shows the SNR distribution of the images in the test set. According to this distribution, images with an SNR lower than 0.4 are defined as low-SNR images, those with an SNR between 0.4 and 0.6 as medium-SNR images, and those with an SNR higher than 0.6 as high-SNR images. The spiral-arm recognition performance is analyzed under three different SNR conditions: high, medium, and low. As can be seen from Fig. 7, the spiral-arm recognition performance is significantly better for high-SNR images.

Fig. 6 SNR distribution of test set images

Fig. 7 Recognition performance under different SNR

5 Conclusion

As the most prominent structural feature of spiral galaxies, spiral arms are closely related in number, shape, and winding tightness to the internal dynamical processes of galaxies, their star-formation histories, and the distribution of dark matter. By observationally and statistically analyzing the distribution of spiral-arm numbers, and combining this with numerical simulations, astronomers can trace the evolutionary trajectories of galaxies and understand the evolutionary laws of spiral structures in different cosmic epochs.

Based on the brightness-distribution characteristics of spiral galaxies, this paper designs a brightness-based bidirectional-fusion Manhattan attention mechanism. Compared with the traditional self-attention mechanism, it can better cope with the brightness-distribution features of spiral-galaxy images, dynamically adjust attention weights according to spatial prior knowledge, and enhance the adaptability of the model. At the same time, for cases in which the brightness center of a spiral galaxy differs from the image center, this study investigates the use of brightness-threshold screening to determine the central coordinates of the spiral galaxy, thereby optimizing the computation of the attention mechanism. By combining the designed attention mechanism with network models such as EfficientNet, a model for recognizing the number of spiral arms in spiral galaxies is constructed. The ablation-experiment results show that the attention mechanism designed in this paper can improve the model' s performance in recognizing the number of spiral arms. In particular, for spiral-galaxy images whose numbers of spiral arms are easily confused and whose data are imbalanced—namely 3, 4, and greater than 4—all evaluation metrics are improved.

References

[1] Semenov V, Tymchyshyn V, Bezguba V, et al. A&C, 2025, 52: 00963

[2] Cao Jie, Xu Tingting, Deng Yuhe, et al. Acta Astronomica Sinica, 2024, 65: 32

[3] Cao J, Xu T T, Deng Y H, et al. ChA&A, 2024, 48: 683

[4] Flores M R, Corral L J, Fierro-Santillán C R, et al. A&C, 2023, 45: 100760

[5] Zhang Jinqu, Ling Han, Du Ping, et al. Acta Astronomica Sinica, 2023, 64: 54

[6] Dong Shuyu, Zhang Jinqu. Acta Astronomica Sinica, 2025, 66: 18

[7] Dong S Y, Zhang J Q. ChA&A, 2025, 49: 536

[8] Lee M W, John Y H, Soo S, et al. A&C, 2025, 52: 100965

[9] Davis D R, Hayes W B. AJ, 2014, 790: 87

[10] Hart R E, Bamford S P, Hayes W B, et al. MNRAS, 2017, 472: 2263

[11] Deutsch T R, Bignone L A, Pedrosa S E. BAAA, 2023, 64: 123

[12] Gong Junyu, Mao Yewai. Acta Astronomica Sinica, 2023, 64: 20

[13] Bahdanau D, Cho K, Bengio Y. 2014, arXiv: 1409.0473

[14] Vaswani A, Shazeer N, Parmar N, et al. 2017, arXiv: 1706.03762

[15] Dosovitskiy A, Beyer L, Kolesnikov A, et al. 2020, arXiv: 2010.11929

[16] Fan Q, Huang H, Chen M, et al. 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle: IEEE, 2024: 5641

[17] Wu D L, Zhang J Q, Li X R, Li H. RAA, 2022, 22: 115011

[18] Willett K W, Lintott C J, Bamford S P, et al. MNRAS,

2013, 435: 2835

[19] Huang G, Liu Z, Van Der Maaten L, et al. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu: IEEE, 2017: 4700

An Attention Mechanism-based Method for Identifying the Number of Spiral Arms in Spiral Galaxies

DU Ping1 DONG Shu-yu2 ZHANG Jin-qu2

(1 School of Building Information, Guangdong Construction Vocational Technology Institute, Qingyuan 511500)
(2 School of Computer Science, South China Normal University, Guangzhou 510631)

Abstract The precise identification of the number of spiral arms in spiral galaxies using deep learning methods plays an important role in tracing the evolutionary trajectories of galaxies and understanding the evolutionary laws of spiral structures in different cosmic periods. However, the features of spiral galaxy images, particularly the distribution characteristics of the spiral center and arm brightness, have not been properly taken into account by existing deep learning models. Therefore, this article proposes a brightness-based bidirectional fusion Manhattan attention mechanism, and combines this attention mechanism with the EfficientNet model for identifying the number of spiral arms in spiral galaxies. An experimental study was carried out using spiral galaxy images selected from the Galaxy Zoo 2 dataset as an example. The results show that the overall accuracy of the EfficientNet model on the test dataset can reach 95.52% by integrating an attention mechanism, with especially good performance in recognizing galaxies with 3, 4, and more spiral arms. Research has shown that well-designed attention mechanisms can reveal the brightness distribution characteristics of spiral galaxies and improve the effectiveness of identifying the number of spiral arms.

Key words galaxies: spiral galaxy, techniques: image processing, methods: data analysis, methods: classification

Note: Figure translations are in progress. See original paper for figures.