2025 5th International Conference on Artificial Intelligence and Industrial Technology Applications (AIITA)

Improved ResNet50 galaxy classification with multi-attention mechanism

Xiaoyu Tong
Institute of Data Management and Applications
China Academy of Industrial Internet
Beijing, China
tongxiaoyu@china-aii.com

Jiehao Chen
Institute of Data Management and Applications
China Academy of Industrial Internet
Beijing, China
*Corresponding author: chenjiehao@china-aii.com

Jie Shi
Institute of Data Management and Applications
China Academy of Industrial Internet
Beijing, China
shijie@china-aii.com

Yuan Jiang
Institute of Data Management and Applications
China Academy of Industrial Internet
Beijing, China
jiangyuan@china-aii.com

***Abstract*—Classification of galaxy morphology has become one of the most important aspects of studying the physical properties of different galaxies. However, the huge amount of galaxy image data generated by large-scale survey programs poses a new challenge to accurate classification. This paper utilizes the Galaxy Zoo 2.0 datasets to classify five types of galaxies while performing data crop, enhancement, and standardization. The ResNet50 neural network algorithm is used, the network characteristics of the algorithm are analyzed in depth, and the channel attention mechanism and spatial attention mechanism are inserted into the residual module, so as to construct the model of CS_ResNet50 improved algorithm to achieve the automatic classification of galaxy morphology. The experimental results show that compared with the ResNet50, CS_ResNet50 improves the model accuracy by 2.85%, precision by 7.25%, recall by 4.23%, and F1-score by 5.14%, which exhibits a strong feature extraction capability and can efficiently identify galaxies with different morphologies, providing a new method for the large-scale morphological classification of galaxies in future large-scale sky survey programs.**

Keywords-Classification of galaxy morphology; channel attention mechanisms; spatial attention mechanisms; CS_ResNet

I. INTRODUCTION

Galaxies are one of the largest celestial systems in the universe, consisting of large numbers of stars, nebulae, star clusters, interstellar gas and dust, and dark matter. The classification of galaxies is a key to the understanding of their formation and evolution. Since the early 20th century, astronomers have been exploring ways to classify galaxies, which has not only improved their understanding of the universe but also advanced theoretical and observational studies in astronomy[1].

The classification of galaxies has advanced with the wide use of large-scale sky surveys and highly accurate observational instruments. The Sloan Digital Sky Survey (SDSS) provides more accurate measurements of the number and mass of galaxies, aiding in the statistical characterization of their distribution[2]. At the same time, analysis of the spectral properties of galaxies has led to a new classification of the dynamical properties and chemical composition of galaxies[3]. Nevertheless, galaxy classification is still a challenging task, and the study of morphological classification of galaxies based on deep learning algorithms has been widely used in astronomy and has become one of the hotspots of research[4]. In deep learning algorithms, raw galaxy data can be fed directly as input to the neural network, and image feature extraction can be completed by the convolutional layer to autonomously uncover complex and efficient higher-order features[5]. In addition, the proposal of the attention mechanism[6] has led to a qualitative leap in the research of neural networks in the field of morphological classification, due to the fact that the attention mechanism makes the classification results more accurate by assigning the appropriate weight values, weakening the useless information and focusing on specific features when the neural network performs feature extraction. Literature [7] proposes a Hybrid Spectral Convolutional Neural Network Attention Mechanism (HybridSN_AM) based on an attention mechanism, to realize high-precision classification of hyperspectral images. Literature [8] proposes a split channel attention mechanism, enhances the diversity of the output features of the global average pooling layer in channel attention, and has better feature extraction and classification capabilities. Literature [9] proposes a few-shot remote scene classification method based on attention mechanism, and the lightweight attention module is introduced into the feature extraction network to improve the classification performance. Literature [10] proposes a model based on a convolutional neural network and attention mechanism to solve the CNN receptive field limitation problem and achieve higher classification accuracy.

In view of this, this paper proposes a ResNet50[11] image classification method incorporating a multi-attention mechanism to improve the performance of the original neural network algorithm in this classification task and to achieve the application of galaxy image classification in the field of astronomy. The method is mainly divided into a data input layer, a model network structure layer, and a data output layer. Data input layer: using the Galaxy Zoo 2.0 datasets [12], set the classification criteria of five galaxy categories, divided into training datasets, test datasets, and validation datasets according to the ratio of 8:1:1, and use image processing methods to operate on the above datasets respectively. Model network structure layer: using a typical residual neural network, ResNet50, introduce a channel attention mechanism[13] and spatial attention mechanism[14] inserted into the residual module to construct the CS_ResNet50 improved algorithm model. Data output layer: set evaluation indices to analyze the experimental effect before and after the algorithm improvement. The experimental data show that CS_ResNet50's results are more accurate in classifying galaxies morphologically.

II. Multiple Attention Mechanisms

A. Channel Attention Mechanism

Channel Attention Mechanism (CAM) models the interdependencies between channels by integrating with various neural network algorithms. This approach captures the relationships among different channels and emphasizes their feature representation within the neural network. As a result, the network can focus more on the channels that are most relevant to the current task, enhancing learning and feature extraction and ultimately improving overall model performance.

The channel attention mechanism described in this paper is illustrated in Fig. 1. Initially, two global pooling operations: average pooling and maximum pooling are applied to the feature map channels to assess the importance of each channel in the input feature map. The pooled results are then reshaped to the size of (1, 1, channels) to prepare them for further processing. A convolution operation is used to analyze the channel features obtained from the global pooling. The features from both pooling results are combined through element-wise addition. To derive the weight distribution of different channels during feature extraction, the sigmoid activation function is applied, mapping the output to the interval [0, 1]. The corresponding calculation formula is presented as in (1):

$$w=\delta(Conv(F_{avg}^{reshape})+Conv(F_{max}^{reshape}))\in R^{1\times 1\times C} \tag{1}$$

The variable w represents the channel attention weight, which indicates the importance of each channel in the feature map. This weight is mapped to specific intervals using the sigmoid activation function δ. $F_{avg}^{reshape}$ refers to shape-adjusted average pooling and $F_{max}^{reshape}$ refers to shape-adjusted maximum pooling.

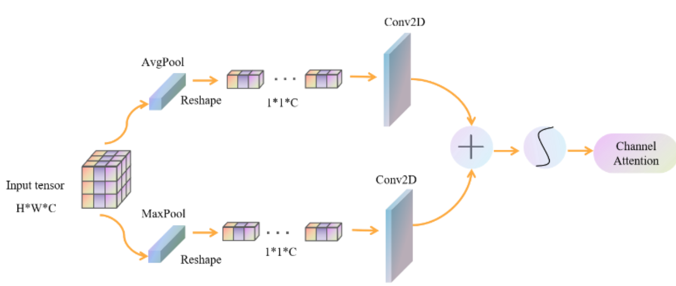


Figure 1. Structure of the channel attention mechanism

B. Spatial Attention Mechanism

Spatial Attention Mechanism (SAM) enhances a model's ability to focus on important regions of input features. Its primary goal is to dynamically adjust the importance of various spatial locations within the input data, allowing the model to emphasize valuable information while suppressing irrelevant details. This helps the neural network perform better in modulation detection tasks and improves its feature representation capabilities.

The spatial attention module presented in this paper is illustrated in Fig. 2. It begins with the creation of a single-channel feature map (H × W × 1) by simultaneously applying two different pooling methods: average pooling and maximum pooling. This approach enables the model to capture richer spatial information through diverse feature aggregation techniques. The results from average and maximum pooling are then combined in the channel dimension to form a two-channel feature map (H × W × 2). Next, a 7×7 convolution operation is applied to these pooled results, resulting in a single-channel spatial attention map (H × W × 1). A sigmoid function is then utilized to generate the attention weight matrix. This matrix helps to identify and enhance the important spatial regions, allowing the spatial information within the network to be utilized more effectively. The calculation formula is provided as in (2):

$$w=\delta(Conv(Stack(F_{avg},F_{max})))\in R^{H\times W\times 1} \tag{2}$$

Where w is the spatial attention weight that identifies and enhances important spatial regions in the input feature map, which are mapped to intervals of the specified range by the sigmod activation function δ. F_{avg} denotes average pooling and F_{max} denotes maximum pooling.

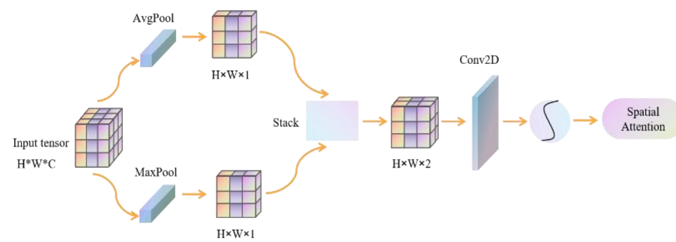


Figure 2. Structure of the spatial attention mechanism

III. Data Preprocessing

A. Data sample selection

The datasets used in this paper originates from the Galaxy Zoo 2.0 project and is publicly available on the official Kaggle website (https://www.kaggle.com/c/Galaxy-Zoo-the-Galaxy-challenge). It includes 61,578 color images along with floating-point numerical data documentation. The image size is 424× 424 × 3. The first column of the data file contains the image ID, which serves to index the corresponding image for that row of data. The following 37 columns consist of floating-point numbers ranging from 0 to 1 (both 0 and 1). These values represent the probability distribution of the galaxies' morphology across 11 directions and 37 categories. Higher values indicate a greater likelihood that the galaxy belongs to a specific morphological category, while lower values suggest a reduced likelihood of being characterized by that category.

B. Classification threshold setting

In conjunction with the Galaxy Zoo 2 decision tree and the official Galaxy Zoo 2 thresholding table, a morphological classification study was carried out using a semi-automatic classification method on the five galaxies for which galaxy classification was achieved in this work. In this classification system, galaxies are categorized using the following numerical designations: 0 represents circular galaxies, 1 is for intermediate galaxies, 2 indicates cigar galaxies, 3 corresponds to edge-on galaxies, and 4 stands for spiral galaxies. The thresholds for each category are detailed in Table I below:

TABLE I. DATA SAMPLE SELECTION RULES

Class	Task	Threshold	Quantities
0	T1.1	Class1.1>= 0.469	8436
	T7.1	Class7.1>= 0.500	
1	T1.1	Class1.1>= 0.469	8069
	T7.2	Class7.2>= 0.500	
2	T1.1	Class1.1>= 0.469	579
	T7.3	Class7.3>= 0.500	
3	T1.2	Class1.2>= 0.430	3903
	T2.1	Class2.1>= 0.602	
4	T1.2	Class1.2>= 0.430	7806
	T2.2	Class2.2>= 0.715	
	T4.1	Class4.1>= 0.619	

The final 28793 galaxy image classification samples are divided into training datasets, test datasets, and validation datasets in a ratio of 8:1:1. Among them, the training datasets contains 23032 image samples, and the test and validation datasets contain 2880 and 2881 image samples respectively. The percentages of the training, test, and validation datasets for each galaxy category are shown in Table II:

TABLE II. DATA SAMPLE NUMBER

Class	Training datasets	Test datasets	Validation datasets
0	6748	844	844
1	6455	807	807
2	463	58	58
3	3122	390	391
4	6244	781	781

C. *datasets processing*

Analysis of the data samples shows that the galaxy images have a large sky background and the target features for which extraction is required are located in the central region of the image. In this paper, the sample data center image is cropped to a size of 224×224×3, which is suitable for the model input, and then standardization. Standardization allows image data to be de-meaned and centered, while data centering matches the distribution of the data and can speed up model convergence.

To increase the diversity of the data, data pre-processing operations are performed on the training datasets before model training to improve the generalization ability and robustness of the model through operations such as cropping, data augmentation, and standardization processing. The type of enhancement used include rotation, flip, brightness and contrast adjustment, etc. Fig. 3 shows the processing flow of the image:

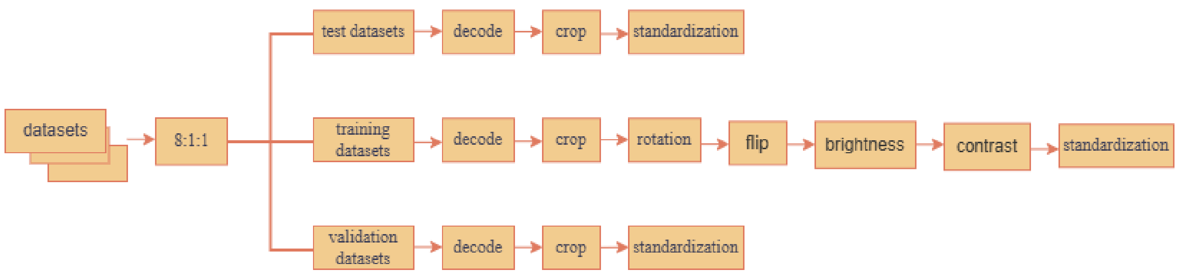


Figure 3. Processing flow of image datasets

IV. NEURAL NETWORK ARCHITECTURE DESIGN

A. *ResNet50*

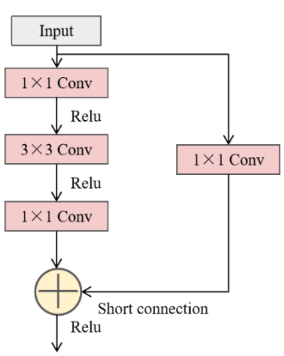


Figure 4. Structure of residual unit

ResNet50 is convolutional neural network that incorporates a residual structure, distinguishing it from traditional convolutional neural networks, as illustrated in Fig. 4. Addresses the lack of performance improvement and the disappearance of gradients as the network gets deeper. The ResNet50 network architecture consists of six components: one input module, four residual modules (comprising 3, 4, 6, and 3 residual units, respectively), and one output module. Together, these components create a deep network structure with a total of 50 layers. The dimensional parameters for each block in the ResNet50 network, along with the output dimensions, are presented in Table III:

TABLE III. RESNET50 NETWORK STRUCTURE

Layer-name	Output size	50-layer
Conv1	112×112	7×7, 64, stride2
Conv2_X	56×56	3×3 max_pool, stride2 $\begin{bmatrix} 1\times1, & 64 \\ 3\times3, & 64 \\ 1\times1, & 256 \end{bmatrix}\times3$
Conv3_X	28×28	$\begin{bmatrix} 1\times1, & 128 \\ 3\times3, & 128 \\ 1\times1, & 512 \end{bmatrix}\times4$
Conv4_X	14×14	$\begin{bmatrix} 1\times1, & 256 \\ 3\times3, & 256 \\ 1\times1, & 1024 \end{bmatrix}\times6$
Conv5_X	7×7	$\begin{bmatrix} 1\times1, & 512 \\ 3\times3, & 512 \\ 1\times1, & 2048 \end{bmatrix}\times3$
	1×1	average_pool, 1000-d fc, softmax

B. CS_ResNet5

In this paper, based on the ResNet50 neural network structure, Channel Attention Mechanism and Spatial Attention Mechanism (CAM and SAM, CS) are added to the residual units of this network, as shown in Fig. 5. CS attention mechanism is combined sequentially, first applying Channel Attention Mechanism and then Spatial Attention Mechanism, to create CS_ResNet50. The network structure is illustrated in Fig. 6. The input to this network is 224 × 224 × 3, and it outputs probability values corresponding to the classification results of five galaxies.

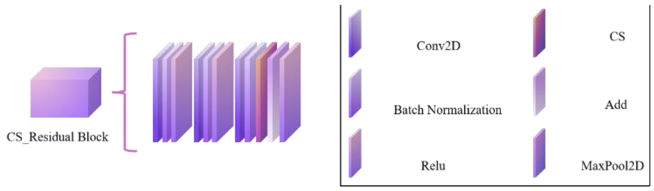


Figure 5. CS_Residual Block Residual structure

As can be seen in Fig. 6, the preprocess imaged galaxy image data is first passed through a convolutional layer with a 7×7 convolutional kernel and a stride of 2, resulting in a 112×112 feature map for low-level feature extraction. Next, a maximum pooling layer with a 3 × 3 convolution kernel and a stride of 2 is introduced to decrease the spatial dimensionality of the feature map. Subsequently, the data is entered sequentially into the improved four-group residual unit, which incorporates the CS module, and the improved residual unit is shown in Fig. 5: The first residual group contains three residual modules and reduces the size of the feature map to 56×56; the second residual group consists of four residual modules and reduces the feature map passing through the first residual group to 28×28; the third residual group contains six residual modules and reduces the feature map passing through the second residual group to 14×14; and the fourth residual group contains three residual modules and reduces the feature map passing through the third residual group to 7×7. After all the residual modules, a global average pooling layer is used instead of the traditional fully connected layer to reduce the number of parameters and prevent model overfitting. Finally, the probabilities of the five galaxy classes are output through the softmax layer.

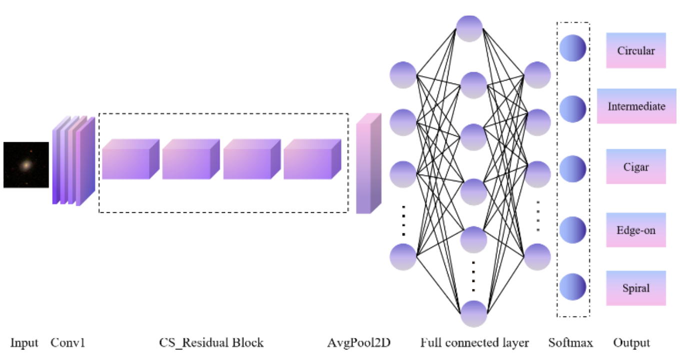


Figure 6. CS_ResNet50 network structure

In the overall structure of the CS_ResNet50 network, both Channel Attention Mechanism and Spatial Attention Mechanism are integrated within each residual unit layer of ResNet50. These modules work in conjunction with the other layers of the network to enhance its overall performance. In CS_ResNet50, the convolutional layer is responsible for extracting layer-level features, while the added attention module enhances the representation of layer-level features by dynamically adjusting the level of attention to the features, enabling the network to more accurately locate and exploit key information in the Galaxy Zoo 2 data.

V. Experimental Results And Analyses

A. Experimental environment

The hardware environment of the computer used is an Intel Gold 5218 processor with 16 cores, 2.3 GHz main frequency, and four GPU graphics cards with 32 GB, Tesla V100 SUPER GPU graphics cards. The software environment is Ubuntu 20.04 64-bit Linux operating system. Based on the deep learning framework TensorFlow 2.15.0 and Python 3.9 as the languages for this development.

B. Evaluation metrics

In this paper, chose to use Confusion Matrix, accuracy, precision, recall, and F1-score to evaluate the performance of ResNet50 and CS_ResNet50 for the validation set in galaxy morphology classification.

The formulas for accuracy, precision, recall, and F1-score are as follows:

$$\text{accuracy}=\frac{TP+TN}{TP+TN+FP+FN} \tag{3}$$

$$\text{precision}=\frac{TP}{TP+FP} \tag{4}$$

$$\text{recall}=\frac{TP}{TP+FN} \tag{5}$$

$$\text{F1-score}=2\times\frac{TP}{2TP+FP+FN} \tag{6}$$

TP (True Positive) refers to the accurate identification of a positive sample as positive.

TN (True Negative) refers to the accurate identification of a negative sample as negative.

FP (False Positive) occurs when a negative sample is incorrectly classified as positive.

FN (False Negative) occurs when a positive sample is incorrectly classified as negative.

C. Model training

To effectively prevent overfitting during the training process, this study employs a regularization technique. Furthermore, it utilizes the control variable method for comparative analysis, ensuring that the hyperparameter settings for model training remain consistent, as detailed in Table IV. This approach enhances the accuracy and comparability of the experimental results. The comparative experiments demonstrate that the improved neural network algorithm, CS_ResNet50, achieves better classification results in the galaxy classification task compared to the original ResNet50 algorithm.

TABLE IV. MODEL TRAINING PARAMETER SETTINGS

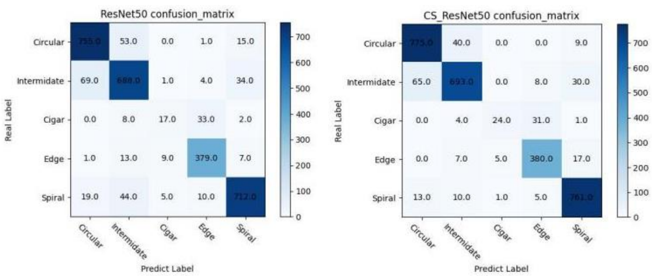
Network	Parameters	Epoch	Learning rate	Batch_size
ResNet50	38203269	100	0.01	128
CS_ResNet50	38601809	100	0.01	128

D. Metric analysis

The experimental data presented in Table V shows that, under identical experimental conditions, the CS_ResNet50 neural network algorithm outperforms the ResNet50 algorithm. Specifically, CS_ResNet50 demonstrates an increase of 2.85% in accuracy, 7.25% in precision, 4.23% in recall, and 5.14% in F1-score. Overall, CS_ResNet50 network exhibits superior performance in identifying and classifying galaxy datasets, including circular, intermediate, cigar, edge-on, and spiral galaxies.

TABLE V. PERFORMANCE COMPARISON

Network	Accuracy	Precision	Recall	F1-score
ResNet50	88.61%	81.83%	77.84%	79.07%
CS_ResNet50	91.46%	89.08%	82.07%	84.21%



(a)ResNet50 (b)CS_ResNet50

Figure 7. Confusion Matrix

As can be seen in Fig. 7 (a) and (b), the CS_ResNet50 neural network algorithm improves the correct classification of circular galaxies by 2.400 %, intermediate galaxies by 0.619 %, cigar galaxies by 12.069 %, edge-on galaxies by 0.256%, and spiral galaxies by 6.274%.

VI. CONCLUSION

In recent years, the use of deep learning algorithms to solve the problem of classifying galaxies, stars, and quasars in celestial sources has been a hot topic in the field of astronomy. In this paper, for ResNet50 neural network algorithm, Channel Attention Mechanism and Spatial Attention Mechanism are introduced, and combined multi-attention mechanism is inserted into the residual module unit, constituting the CS_ResNet50, which optimizes the performance of the network and achieves the correct classification of the data samples of circular galaxies, intermediate galaxies, cigar galaxies, edge-on galaxies and spiral galaxies. The experimental results show that CS_ResNet50 improves the model accuracy by 2.85%, precision by 7.25%, recall by 4.23%, and F1-score by 5.14% compared to the ResNet50 neural network algorithm. In addition, the proportion of correct classifications for each galaxy class is also improved in terms of galaxy classes. Therefore, the improved algorithmic model compensates for the shortcomings of the original model in terms of classification precision and accuracy and is more effectively used for the morphological classification of galaxies, which can provide strong support for the morphological classification of galaxies.

REFERENCES

[1] Owens, E, A, et al. Using oblique decision trees for the morphological classification of galaxies[J]. Monthly Notices of the Royal Astronomical Society. 1996,281:153-157.

[2] York D G, Adelman J, Anderson Jr J E, et al. The sloan digital sky survey: Technical summary[J]. The Astronomical Journal, 2000, 120(3): 1579.

[3] Pacifici C, Iyer K G, Mobasher B, et al. The art of measuring physical parameters in galaxies: a critical assessment of spectral energy distribution fitting techniques[J]. The Astrophysical Journal, 2023, 944(2): 141.

[4] Lanusse F. The Dawes Review 10: The impact of deep learning for the analysis of galaxy surveys[J]. Publications of the Astronomical Society of Australia, 2023, 40: e001.

[5] YANG Y, WEN Z L, XIA J Q. Star-galaxy separation by the residual neural network algorithm[J]. Journal of Beijing Normal University (Natural Science), 2021, 57(4): 450-457. DOI: 10.12202/j.0476-0301.2021106.

[6] Niu Z, Zhong G, Yu H. A review on the attention mechanism of deep learning[J]. Neurocomputing, 2021, 452: 48-62.

[7] LIU Y J, LIU Y D, YAN Z, etc. Classification of Hybrid Convolution Hyperspectral Images Based on Attention Mechanism[J]. Spectroscopy and Spectral Analysis, 2024,44(10):2916-2922.

[8] CHU Y Z, SHI Y J， ZHANG X F， LIU H. Image classification algorithm based on split channel attention network[J]. Chinese Journal of Engineering, Vol. 46, No. 10: 1856−1863, October 2024.

[9] ZHANG D N, ZHAO H J, LU Y Y, etc. Few-Shot Scene Classification with Attention Mechanism in Remote Sensing[J]. Computer Engineering and Applications, 2024,60(04):173-182.

[10] GAO Y P, YAN W H, PAN X. Hyperspectral image classification based on convolutional neural network and attention mechanism[J]. Journal of Optoelectronics· Laser, 2024, 35(5): 483-489.

[11] He K, Zhang X, Ren S, et al. Identity mappings in deep residual networks[C]//Computer Vision-ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part IV 14. Springer International Publishing, 2016: 630-645.

[12] Willett K W, Lintott C J, Bamford S P, et al. Galaxy Zoo 2: detailed morphological classifications for 304 122 galaxies from the Sloan Digital Sky Survey[J]. Monthly Notices of the Royal Astronomical Society, 2013, 435(4): 2835-2860.

[13] Mekruksavanich S, Jitpattanakul A. Hybrid convolution neural network with channel attention mechanism for sensor-based human activity recognition[J]. Scientific Reports, 2023, 13(1): 12067.

[14] Ghosh S, Chatterjee A. T-Fusion Net: A Novel Deep Neural Network Augmented with Multiple Localizations based Spatial Attention Mechanisms for Covid-19 Detection[C]//International Conference on Computer Vision and Image Processing. Cham: Springer Nature Switzerland, 2023: 213-224.