Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Chrysanthemum Classification Method Integrating Deep Visual Features from both the Front and Back Sides

Yifan Chen
School of Computer and Electronic Information /School of Artificial Intelligence, Nanjing Normal University
Xichen Yang
xichen_yang@njnu.edu.cn

School of Computer and Electronic Information /School of Artificial Intelligence, Nanjing Normal University
Hui Yan
National and Local Collaborative Engineering Center of Chinese Medicinal Resources Industrialization and Formulae Innovative Medicine
Jia Liu
College of Artificial Intelligence and Information Technology, Nanjing University of Chinese Medicine
Jian Jiang
School of Computer and Electronic Information /School of Artificial Intelligence, Nanjing Normal University
Zhongyuan Mao
School of Computer and Electronic Information /School of Artificial Intelligence, Nanjing Normal University
Tianshu Wang
College of Artificial Intelligence and Information Technology, Nanjing University of Chinese Medicine

Method Article

Keywords: Chrysanthemum classification, Two-stream network, Visual information, Feature fusion

Posted Date: February 13th, 2024

DOI: https://doi.org/10.21203/rs.3.rs-3940219/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Chrysanthemum Classification Method Integrating Deep Visual Features from both the Front and Back Sides

Yifan Chen[1†], Xichen Yang[1*†], Hui Yan[2,3*], Jia Liu[4], Jian Jiang[1], Zhongyuan Mao[1], Tianshu Wang[4,5]

[1*]School of Computer and Electronic Information /School of Artificial Intelligence, Nanjing Normal University, Nanjing, 210046, Jiangsu Province, China.

[2]National and Local Collaborative Engineering Center of Chinese Medicinal Resources Industrialization and Formulae Innovative Medicine, China.

[3]Jiangsu Collaborative Innovation Center of Chinese Medicinal Resources Industrialization, Nanjing University of Chinese Medicine, Nanjing, 210023, Jiangsu Province, China.

[4]College of Artificial Intelligence and Information Technology, Nanjing University of Chinese Medicine, Nanjing,210023, Jiangsu Province, China.

[5]Jiangsu Province Engineering Research Center of TCM Intelligence Health Service, Nanjing University of Chinese Medicine, Nanjing,Jiangsu Province, China.

*Corresponding author(s). E-mail(s): xichen_yang@njnu.edu.cn; yanhui@njucm.edu.cn;
Contributing authors: yifan_chen@nnu.edu.cn; 20230843@njucm.edu.cn; jianjiang@nnu.edu.cn; 19200205@njnu.edu.cn; wangtianshu@njucm.edu.cn;
†These authors contributed equally to this work.

Abstract

Chrysanthemum morifolium Ramat(hereinafter referred to as Chrysanthemum) is one of the most beloved and economically valuable Chinese herbal crops, which

contains abundant medicinal ingredients and wide application prospects. Therefore, identifying the classification and origin of Chrysanthemum is important for producers, consumers, and market regulators. The existing Chrysanthemum classification methods mostly rely on visual subjective identification, are time-consuming, and always need high equipment costs. A novel method is proposed to accurately identify the Chrysanthemum classification in a swift, non-invasive, and non-contact way. The proposed method is based on the fusion of deep visual features of both the front and back sides. Firstly, the different Chrysanthemums images are collected and labeled with origins and classifications. Secondly, the background area with less available information is removed by image preprocessing. Thirdly, a two-stream feature extraction network is designed with two inputs which are the preprocessed front and back Chrysanthemum images. Meanwhile, the incorporation of single-path residual connections and cross-path residual connections is employed to extend the receptive field of the network and fully fusion the features from both the front and back sides. Experimental results demonstrate the proposed method outperforms existing methods, which gain higher accuracy and stability. The proposed method presents an effective and dependable solution for identifying the classification and origin of Chrysanthemum. Code and data are available at https://github.com/dart-into/CCMIFB.

Keywords: Chrysanthemum classification, Two-stream network, Visual information, Feature fusion

1 Introduction

Chrysanthemum morifolium Ramat(hereinafter referred to as Chrysanthemum, traditional iconic flowers in China, boasts a wide variety of species, versatile applications, and a long-standing cultivation history. Chrysanthemums exhibit remarkable aesthetic appeal and hold substantial economic as well as medicinal value. They abound in flavonoids[1] , volatile oils[2], chlorogenic acid[3], sesquiterpenes[4], triterpenes, and amino acids, delivering advantageous health attributes like anti-inflammatory[5], antimicrobial, antioxidant[6], anti-HIV, and anticancer effects[7]. Chrysanthemums have found wide applications in food, tea, ornamentation, and pharmaceuticals. Particularly, Chrysanthemum tea is highly cherished for its health benefits like heat-clearing, digestion-enhancing, liver-nourishing, and vision-boosting effects[8]. Nevertheless, their quality and pricing are contingent upon their place of origin. Identifying the classification and origin of Chrysanthemum rapidly and precisely is important for producers, consumers, and market regulators.

Traditional methods of origin identification mainly comprise plant phenotype analysis and physicochemical analysis. Plant phenotype analysis heavily relies on human intervention, involving expertly trained assessment teams to primarily discern the characteristics of Chrysanthemums, including shape, color, and odor. The huge group of China's traditional Chrysanthemum varieties, coupled with the yearly cultivation of numerous new breeds, contribute to an immensely diverse array of species. Additionally, Chrysanthemums display a rich array of colors and intricate floral structures. Its

inflorescences are diverse in shape, with variable proportions of tubular and ligulate flower composition, especially characterized by an increase in the number of ligulate flowers. Many Chrysanthemum varieties share a high degree of similarity, making manual identification methods time-consuming, prone to errors, and challenging for large-scale variety classification and identification tasks. In the realm of physicochemical analysis of plants, researchers have introduced various analytical instruments and conducted extensive studies. Techniques such as GC-MS, electronic nose[9], FT-IR[10], ICP-MS[11], LC×LC-Q-TOF/MS[12], in conjunction with chemometrics, have been utilized for the identification of Chrysanthemum origins. These methods primarily ascertain the geographical origin and category origin of Chrysanthemums by detecting the content of effective components, providing advantages of excellent repeatability and high sensitivity. Nevertheless, these analytical methods face challenges such as intricate sample preparation, prolonged analysis duration, and overreliance on costly equipment such as infrared spectrometers and electronic nose devices. Therefore, the urgency to develop a swift, non-invasive, and non-contact method for Chrysanthemum classification becomes especially pronounced.

In recent years, ample research has demonstrated that the future progress of plant phenotype analysis relies on the utilization of computer image processing and data-driven machine learning (ML) methods. This method has surpassed the limitations of manually gathering plant phenotype data, which offers a robust avenue for plant phenomics research. Initially, image classification frequently utilizes shallow machine learning techniques such as K-Nearest Neighbors (KNN) and Support Vector Machine (SVM)[13]. KNN establishes a distance metric to gauge the distance between test samples and those in the training set, while SVM learns weight matrices per class via training, substituting the distance used in KNN. Ever since the inception of deep learning, substantial advancements have occurred in image classification research. Within deep learning, image features are autonomously extracted from datasets, thereby freeing the parameters from human-imposed constraints. Subsequently, the inception of the ImageNet dataset[14] and the continuous hosting of seven ImageNet Large Scale Visual Recognition Challenges (ILSVRC) have engendered several iconic network models such as AlexNet[15], VGGNet[16], GoogLeNet[17], ResNet[18], among others. As research delved deeper, researchers began tackling more challenging tasks such as fine-grained recognition. Given the diverse morphological variations in chrysanthemums, identifying their varieties falls within the realm of fine-grained recognition, offering valuable research data for the fine-grained identification of plant species. Attention[19] mechanisms have found extensive application across various computer vision tasks. Certain methodologies leverage Vision Transformer(ViT)[20] for feature extraction, yielding highly promising outcomes.

Computer vision technology has begun playing an increasingly vital role in botanical morphology studies. Analyzing and processing plant image data facilitates the automatic measurement of plant morphological features, including leaf area, presenting a swift, efficient, and precise method for botanical morphological research. Some researchers have devised learning-enhanced methodologies integrating the Inception-v4 convolutional neural network for grading and fraud detection of saffron images taken with smartphones[21]. Additionally, Researchers utilize computer vision techniques

to detect pests[22], plant diseases[23], maturity[24] in citrus fruits and more. These endeavors illustrate the extensive prospects of computer vision technology in botanical morphological studies. Recent research has aimed to utilize image feature extraction methods for identifying and detecting Chrysanthemum varieties. Some researchers undertook preliminary variety identification by extracting Gray Level Co-occurrence Matrix(GLCM)[25] textures from images of 20 ornamental Chrysanthemum varieties. Additionally, other researchers utilized hyperspectral imaging to measure spectral reflectance values from various parts of Chrysanthemum ray florets and analyzed their correlation with the measured pigment content. Moreover, researchers extracted Local Binary Pattern (LBP)[26] texture features from unfolded images of 24 Chrysanthemum varieties. Yet, these methods were limited to recognizing the overall inflorescence shape and ray floret patterns of Chrysanthemums, lacking the capability for precise identification of Chrysanthemum varieties.

To address the aforementioned issues, this paper presents a Chrysanthemum classification via the fusion of deep visual features of both the front and back sides. This method employs a deep neural network to extract the color, texture, and shape features of Chrysanthemums, facilitating the classification of Chrysanthemum species. The key steps are as follows. Firstly, the different Chrysanthemum images are collected and labeled with origins and classifications. Secondly, the background area with less available information is removed by image preprocessing. Thirdly, a two-stream feature extraction network is designed with two inputs which are the preprocessed front and back Chrysanthemum images. Meanwhile, the incorporation of single-path residual connections and cross-path residual connections is employed to extend the receptive field of the network and fully fusion the features from both the front and back sides.

The primary contributions of this paper include:

(1) Proposing subjective screening criteria for Chrysanthemum images, excluding non-compliant ones, thereby providing a clean and standardized dataset for classification.

(3) Presenting a strategy for integrating deep features from both sides through inter-layer and inter-path interactions, facilitating a comprehensive fusion of front and back Chrysanthemum features.

(3) Presenting a strategy for integrating deep features from both sides through inter-layer and inter-path interactions, facilitating comprehensive fusion of front and back chrysanthemum features.

The subsequent sections of this paper are organized as follows: Section 2 delves into relevant work on image classification and Chrysanthemum species recognition. Section 3 elaborates on the structure and specifics of the proposed network. Section 4 delineates the results of performance evaluation experiments. Finally, Section 5 concludes the paper.

2 Related work

2.1 Image Classification Based on Visual Features

Image classification is a fundamental challenge in computer vision, aiming to categorize images into distinct classes. Traditional machine learning methods have been utilized for image classification over recent decades. Yet, with the evolution of deep learning, deep neural networks have emerged as the forefront method for image classification. ResNet (Residual Network), introduced by He et al.[18], enhances accuracy by considerably increasing depth. Its internal residual blocks use skip connections, alleviating the vanishing gradient issue in deep neural networks. Lin et al.[27] introduced BCNN (Bilinear CNN), which utilizes bilinear pooling to grasp pixel relationships among images for classification. It captures global feature information by computing the outer product of two CNN feature maps, making it a highly representative model for fine-grained weakly supervised learning. Researchers from Stanford University and Facebook AI Research introduced RegNet[28] which is a novel convolutional neural network. It achieves heightened accuracy and reduced parameter count through automated network structure search. EfficientNet[29], published by Google, improves model performance through balanced scaling in depth, width, and resolution. With the advent of attention mechanisms, numerous researchers have implemented them in image classification tasks. SENet[30] (Squeeze-and-Excitation Network), introduced by Hu et al., dynamically adjusts feature map weights using attention mechanisms to enhance image classification performance. CBAM[31] (Convolutional Block Attention Module), introduced by Woo et al., combines channel and spatial attention to optimize the model's attention on various image regions, thereby improving image classification performance. Swin-T[32], proposed by researchers from The Chinese University of Hong Kong and Microsoft Research Asia, achieves increased accuracy and decreased computational complexity through multi-level partitioned attention mechanisms. TinyViT[33], introduced by researchers from Microsoft, is a novel compact ViT. It transfers knowledge from large pre-trained models to smaller ones using swift pre-training distillation methods, allowing smaller models to leverage abundant pre-training data. ConvNeXtV2[34], an extension derived from the ConvNeXt[35] architecture inspired by MAE[36], introduces a new Global Response Normalization (GRN) layer to enhance competitive feature expression among channels within the original ConvNeXt modules, thereby better capturing discriminative channel features.

Presently, within the realm of Chrysanthemum research, deep learning methods are progressively emerging. Fu et al.[37] combined hyperspectral imaging technology with chemometrics to explore and apply the discrimination of Hangbaiju's origins. Yuan et al.[38] employed convolutional neural networks for Chrysanthemum flower identification, achieving an identification rate of approximately 95%. Nonetheless, it categorized only 5 flower types and could not accurately discern varieties. Following that, they introduced a Chrysanthemum image phenotype classification framework based on transfer learning and bilinear convolutional neural networks[39]. Utilizing a symmetrical VGG16 network as a feature extractor, they eventually fed global features into the classification layer for sorting. Liu et al.[40]. captured 14,000 images

of 103 large Chrysanthemum varieties using an image acquisition device. Leveraging the concept of transfer learning, they established a recognition model for these varieties based on VGG16, GoogLeNet, and ResNet50 deep convolutional neural networks. However, the clustering and visualization of extracted deep features did not manifest distinct distribution patterns. Wang et al.[41] integrated AP clustering analysis with deep features, proposing a multi-information model based on deep learning for the identification and classification of large-flowered Chrysanthemums. Building upon the traditional VGG16 convolutional neural network, Huang et al.[42] introduced an enhanced multi-scale, multi-parallel convolutional neural network termed VGG-Inception. Utilizing parallel network structures and global pooling layers maintains model depth and augments network width while reducing parameters to just 9% of the original VGG16. The utilization of auxiliary classifiers mitigates gradient vanishing, thereby enhancing the model's ability to generalize.

2.2 Deep Multi-Path Networks

Deep Multi-Path Networks represent intricate neural network structures employing multiple parallel processing pathways to concurrently handle diverse inputs or feature channels, striving to comprehensively capture image information and bolster processing efficiency. This architecture permits diverse pathways to specialize in distinct feature extraction or tasks. By amalgamating the outputs from these pathways, it generates more precise or comprehensive outcomes, exhibiting significant relevance across multiple domains of image processing. Ma et al.[43] introduced a dual-stream network based on wavelet transform to address color deviations and blurry details in underwater images. Pan et al.[44] introduced a method to enhance synthesized view quality based on a dual-stream attention network (TSAN). The global information extraction stream learns contextual information, while the local information extraction stream extracts texture details from rendered images. The Multi-Scale Residual Attention Block (MSRAB) effectively detects features of varying scales and optimizes them by considering spatial interdependencies. Liu et al.[45] proposed a driver fatigue detection method utilizing a dual-stream network model integrating various facial features. It extracts static features from partial facial images and dynamic features from partial facial optical flow, feeding them into a dual-stream neural network for feature fusion and classification, showcasing excellent performance. Zhou et al.[46] harnessed EfficientNet's efficient feature extraction abilities to separately extract spatial and temporal features of consecutive video frames from spatial and temporal flows. They then employed a multi-head attention mechanism to capture pivotal action details, facilitating action recognition using the amalgamated features. Wang et al.[47] introduced a global-local dual-path architecture for multi-scale representation, aimed at resolving classification performance constraints due to extensive feature variations in remote sensing images. The Correlation-Driven Joint Bone-Flow Graph Convolutional Network (CD-JBF-GCN), devised by Tu et al.[48], delves into the motion transmission amid joint and bone flows, facilitating more discernible feature representations in both streams. This network demonstrates cutting-edge performance in semi-supervised skeleton action recognition. Zheng et al.[49] introduce a novel Cross-Attention and Cross-Scale Fusion Network (CASF-Net) that maximizes the potential of dual-path

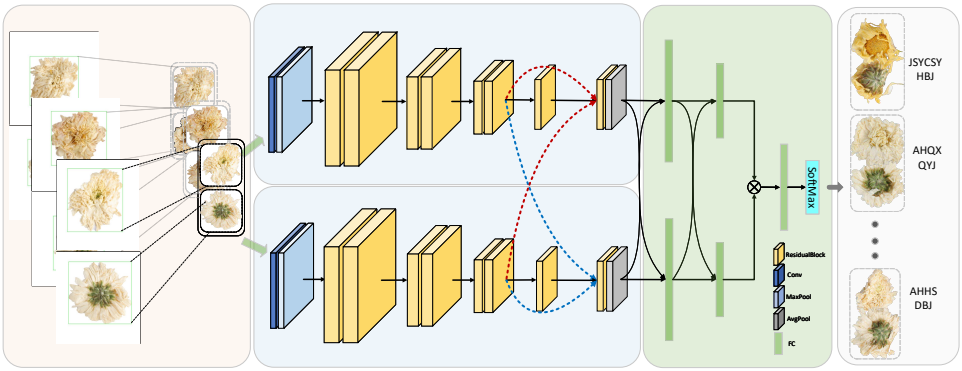


Fig. 1 The architecture of the proposed method.

networks and fully integrates coarse-grained and fine-grained feature representations. The designed dual-branch encoder focuses on modeling non-local dependencies and multi-scale contexts, markedly improving semantic segmentation quality. Similarly, Xie et al.[50] introduce the Context-Aware Network with Dual-Stream Pyramid (CANet) tailored for medical image segmentation. Through multiple resolution input versions and multi-scale convolutional units, CANet adeptly captures diverse hierarchical multi-scale complementary features in medical images. SETNet, presented by Ma et al.[51], stands as a dual-stream convolutional network for no-reference image quality assessment. While the image stream attends to the entire image content, the saliency stream explicitly guides the network in learning spatially significant features more appealing to human perception. Leveraging spatial and channel attention modules refines features and amalgamates multi-level features to predict image quality scores. Drawing inspiration from the human visual system, Chen et al.[52] present a dual-path convolutional neural network for blind image quality assessment tasks. Emulating the two pathways of the human eye, the model extracts image content and global shape features, integrating multi-scale features to enhance assessment performance.

3 Method

To achieve efficient identification of chrysanthemum species, the specific workflow as depicted in Figure 1 is outlined in this paper. It primarily encompasses three components: 1) data preprocessing, 2) feature extraction, and 3) feature fusion.

3.1 Data Preprocessing

Before network input, all Chrysanthemum images undergo a series of steps. First, Canny edge detection method is employed to detect their edges. Next, a minimal square encompassing all edges is used to crop the image, removing the white background devoid of Chrysanthemum information. This method maximizes Chrysanthemum feature extraction. Subsequently, all dataset samples are standardized to 224x224 pixels to mitigate accuracy reduction arising from differing image scales.

Then, all images are randomly rotated left or right by 15° to diversify the training set and mitigate overfitting risks. Lastly, all dataset samples are normalized to expedite model convergence and bolster its generalization. Figure 1 depicts the preprocessing of Chrysanthemum images in the dataset, highlighting edge detection (indicated by the green box), cropping, and resizing for subsequent network input.

3.2 Feature Extraction

Acknowledging the distinct features between the front and back of various Chrysanthemum species, the method proposed in this study employs a two-stream convolutional neural network model, utilizing an enhanced ResNet-18 model for both streams. Inputs consist of images capturing the front and back of Chrysanthemums, tailored to extract crucial features specific to each side, considering the unique structure and morphology of the flowers. A traditional ResNet-18 structure comprises 17 convolutional layers, 1 Maxpooling layer, 1 Avgpooling layer, 1 fully connected layer, and 1 Softmax layer. In contrast to this, our proposed network extends the end fully connected layer to three. The network architecture primarily consists of shallow feature extraction, encompassing the initial 13 convolutional layers, and deep feature fusion, comprising the final 4 convolutional layers and 2 fully connected layers. In the end, the upper and lower stream features are concatenated and merged through a single fully connected layer, as depicted in Table 1 outlining the network structure parameters.

Within Table 1, "Block name" represents module names, while "Layer name" denotes operation names—'conv' for convolution, 'BN' for batch normalization, 'relu' for activation function, and 'Maxpool' for max-pooling operation. 'Parameters' refer to the convolutional kernel size and output channel quantity. Starting with 224*224*3 chrysanthemum phenotype image data, a 7*7 convolutional layer with a stride of 2 and padding of 3 produces output data sized 112*112*64. Subsequent max-pooling, using a 3*3 kernel, a stride of 2, and padding of 1, results in data sized 56*56*64, halving the feature's dimensions without altering the channel count. Following this, Conv2, with a stride of 2 and padding of 1, maintains the data's size and channel count. Then, Conv3, Conv4, and Conv5 double the channel count while halving the output data size, resulting in 7*7*512 output dimensions. A final global average pooling layer outputs data sized 1*1*512. The first two fully connected layers transform data to 128 dimensions and 32 dimensions, respectively. These two 32-dimensional vectors concatenate into a 64-dimensional vector, fed into the last fully connected layer, yielding the ultimate feature vector.

3.3 Feature Fusion

Inspired by residual networks, this paper introduces cross-path residual connections to simultaneously extract features and enhance the network's field of view, allowing better integration of both positive and negative chrysanthemum features.

As illustrated in Figure 2.(a), the interaction mode of residual connections between front and back layers is a traditional single-path ResNet network. Here, x_l represents the output of the previous layer, $h(x_l)$ denotes direct mapping (the left line in (a)), while $F(x_l, W_l)$ signifies residual (the convolutional part on the right side in (a)). The

Table 1 The parameters of the network.

Block name	Layer name	parameters	Layer name	parameters
Conv1	Conv	$7 \times 7, 64$	Conv	$7 \times 7, 64$
	Maxpool	3×3	Maxpool	3×3
Conv2	Conv	$\begin{bmatrix} 3 \times 3, 64 \\ 3 \times 3, 64 \end{bmatrix} \times 2$	Conv	$\begin{bmatrix} 3 \times 3, 64 \\ 3 \times 3, 64 \end{bmatrix} \times 2$
	Bn+relu		Bn+relu	
Conv3	Conv	$\begin{bmatrix} 3 \times 3, 128 \\ 3 \times 3, 128 \end{bmatrix} \times 2$	Conv	$\begin{bmatrix} 3 \times 3, 128 \\ 3 \times 3, 128 \end{bmatrix} \times 2$
	Bn+relu		Bn+relu	
Conv4	Conv	$\begin{bmatrix} 3 \times 3, 256 \\ 3 \times 3, 256 \end{bmatrix} \times 2$	Conv	$\begin{bmatrix} 3 \times 3, 256 \\ 3 \times 3, 256 \end{bmatrix} \times 2$
	Bn+relu		Bn+relu	
Conv5	Conv	$\begin{bmatrix} 3 \times 3, 512 \\ 3 \times 3, 512 \end{bmatrix} \times 2$	Conv	$\begin{bmatrix} 3 \times 3, 512 \\ 3 \times 3, 512 \end{bmatrix} \times 2$
	Bn+relu		Bn+relu	
	Avgpool	7×7	Avgpool	7×7
FC Block	FC1:output nodes128		FC2:output nodes128	
	FC3:output nodes32		FC4:output nodes32	
	FC5:output nodes18			

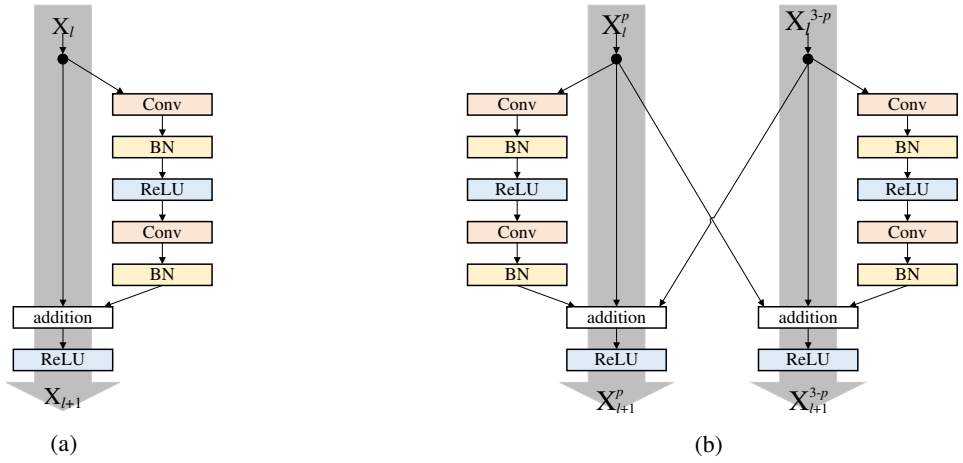


Fig. 2 (a) The traditional residual connections. (b) The cross-path residual connections proposed in this paper.

formula for the output of this layer x_{l+1} is as follows.

$$x_{l+1} = h(x_l) + F(x_l, W_l) \tag{1}$$

Figure 2.(b) illustrates the interaction mode of residual connections between the upper and lower paths within the dual-path convolutional neural network proposed in this paper. Here, $x_l^p (p \subset 1,2)$signifies the output of the upper or lower path network in the previous layer. When p equals 1, x_l^p represents the output of the upper path network processing the front image in the prior layer, and when p equals 2, x_l^p corresponds to the output of the lower path network processing the back image in the previous layer. h denotes an identity mapping, Re indicates the joint residual of the upper and lower paths in the prior layer. This layer's output is denoted as x_{l+1}^p, F represents the convolutional layer's transformation, and the respective formula is delineated below:

$$Re = h(x_l^p) + h(x_l^{3-p}) \tag{2}$$

$$x_{l+1} = Re + F(x_l^p, W_l^p) \tag{3}$$

During residual connections between preceding and subsequent layers, a cross-path residual connection was applied, summing up the three outputs. The use of single-path and cross-path residual connections expands the network's field of view, aiding in capturing diverse scale and abstract features within images. It facilitates direct information exchange between layers within the same path and between different paths, enabling deeper layers to benefit from shallower layers, and ultimately allowing a comprehensive fusion of features from both the front and back sides.

Similar strategies are employed in the fully connected layers to further amalgamate features from both the front and back sides:

$$y_{l+1}^p = G(y_l^p + \frac{y_l^{3-p}}{3}) \tag{4}$$

$y_l^p (p \subset 1,2)$signifies the output of the upper or lower path network in the preceding layer. For p equal to 1, y_l^p represents the upper path network's output, and for p equal to 2, it corresponds to the lower path network's output. Meanwhile, y_l^{p+1} denotes the network's output at this layer, with G representing the transformation in the fully connected layer. Unlike the convolutional layers, this process involves a weighted sum of the upper and lower path features, with a ratio of 1:3. This weighting ratio's validation in subsequent experimental segments underscores its crucial role in the model's performance.

This approach accounts for the potential similarity in front features but dissimilarity in back features among different Chrysanthemum species, or vice versa. This enhances the comprehensive integration of both front and back features, leading to more accurate identification of Chrysanthemum types.

3.4 Model Training

In the training section of the network model, the dataset was divided into training, validation, and test sets at a ratio of 6:2:2. Staged training methods were utilized

to enhance efficacy. Pretraining parameters from the ImageNet dataset were initially transferred, serving as the initial parameters for the first 13 layers of the two-stream network to expedite convergence and enhance effectiveness. These parameters were then frozen for preliminary training, facilitating the rapid acquisition of general features by the model. This phase comprised 20 iterations, a learning rate of 0.001, and a batch size of 24. Subsequently, the parameters for shallow feature extraction were unfrozen for model fine-tuning, involving 30 iterations, a learning rate of 0.0001, and a batch size of 24. The Adam optimizer was used to optimize the model, employing cross-entropy loss as the selected loss function. After each iteration, validation was conducted on the validation set, saving models that showed superior results to achieve the network's optimal solution. To mitigate experimental errors caused by random sampling, the train-test prediction process was repeated 50 times, and the average accuracy was computed.

4 Experimental Results

4.1 Database and Evaluation Metrics

The dataset used in this experiment comprises Chrysanthemums from various regions and categories. Samples of Chrysanthemum morifolium 'Hangbaiju' were collected from Zhoukou, Henan; Sheyang, Yancheng, Jiangsu; Xifeng, Guizhou; Shimen and Wuyi Baimuxiang, Tongxiang, Zhejiang; Julu Hebei and Suizhou, Hubei. Samples of Chrysanthemum morifolium 'Gongju' were obtained from Huangshan, Anhui, and Shangqiu, Henan. Other varieties included Chrysanthemum morifolium'Boju' from Bozhou, Anhui, Chrysanthemum morifolium 'Qiyueju' from Shexian, Anhui, Chrysanthemum morifolium 'Dabanju' and Chrysanthemum morifolium 'Jinsiju' from Huangshan, Anhui, Chrysanthemum morifolium 'Taoju' from Jiujiang, Jiangxi, Chrysanthemum morifolium 'Xiangju' from Xiangshui, Yancheng, Jiangsu, and Chrysanthemum morifolium 'Huaiju' and Chrysanthemum morifolium 'Qibaiju' from Jiaozuo, Henan. The dataset comprises 18 Chrysanthemum species, with approximately 100 images of each Chrysanthemum's front and back sides, resulting in a total of around 3600 images. Figure 3 illustrates schematic images of chrysanthemums from different regions.

During image collection, four criteria were applied to select images that met the standards: While capturing images of the front and back of Chrysanthemums, attention was paid to angle selection. Front images were ensured to exclude receptacles, while the focal point of back images was positioned between the receptacle and calyx, typically with a larger receptacle portion. It is crucial to maintain the integrity of the Chrysanthemum's receptacle and petals, as the absence of the receptacle may lead to irregular shapes, and missing petals may result in occlusion issues. The color of the flowers should resemble the majority within their respective categories, avoiding excessively dark or bright shades. Additionally, the size of the flowers should match the majority of flowers in their respective categories, avoiding situations where they are too large or too small. This study employs four metrics—Accuracy, Recall, F1 Score, and Standard Deviation (STD) between categories—to assess the performance of Chrysanthemum image classification methods. Accuracy indicates the proportion of correctly classified samples by the classifier, while Recall measures the classifier's

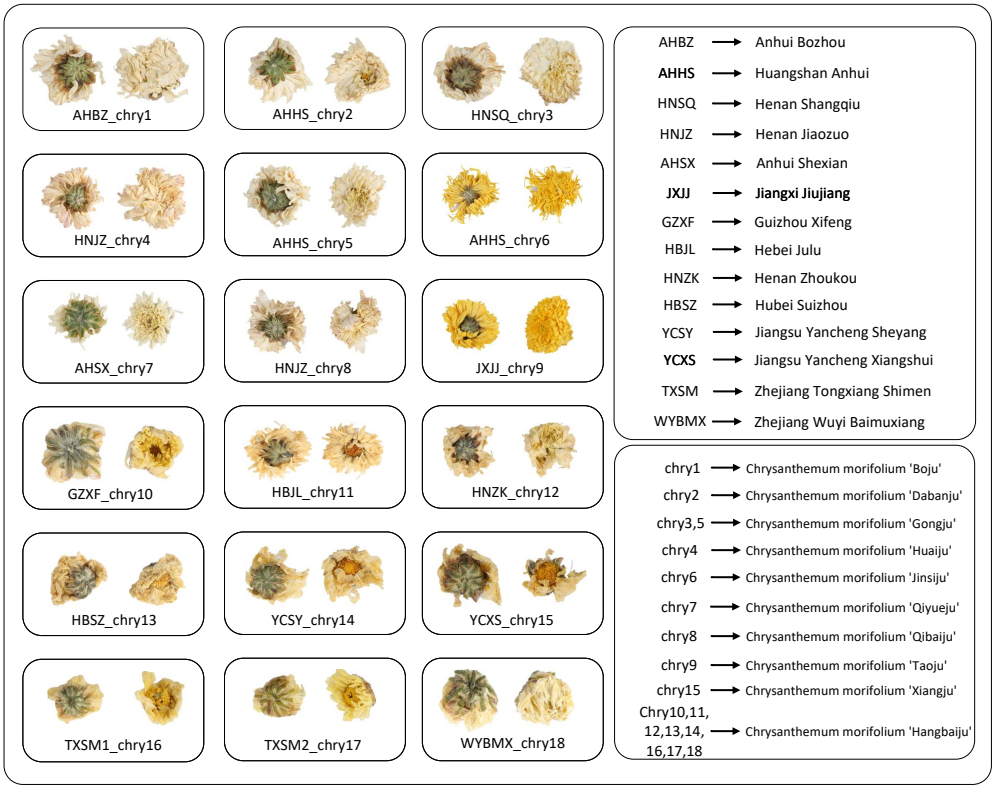


Fig. 3 The illustration of Chrysanthemums.

recognition capability for each category. The F1 Score, a combination of Precision and Recall, offers an overall evaluation of the classifier's performance. The Standard Deviation between categories reflects the method's diversity and balance across various Chrysanthemum categories. These metrics provide multiple perspectives that facilitate a comprehensive understanding of the classifier's performance in Chrysanthemum classification tasks.

4.2 Effectiveness of backbone selection

This study conducted a series of comparative experiments to validate the effectiveness of employing the ResNet network as the backbone for Chrysanthemum classification. CNN, Alex, VGG, GoogLeNet, and DenseNet were separately chosen as the backbone for the two-stream Chrysanthemums classification network. The detailed experimental results are provided in Table 2.

The data from Table 2 indicates that using ResNet as the backbone yielded the best performance across all evaluation metrics. Our model (ours) achieved the highest levels of accuracy, recall, and F1 score, reaching 93.8%, 93.5%, and 93.4% respectively. These rankings place our model at the forefront among all models, highlighting

Table 2 Performance comparisons of different networks as backbone

Backbone	Acc(%)	Recall(%)	F1(%)	STD
TwoStreamAlex	89.1	88.0	88.1	7.22
TwoStreamVGG	84.9	83.4	83.7	9.44
TwoStreamGoogLeNet	90.9	89.9	90.0	7.58
TwoStreamCNN	88.6	87.9	88.0	7.24
TwoStreamDenseNet	93.0	92.1	92.2	5.93
Ours	**93.8**	93.5	93.4	5.20

its superior classification performance. Our model showed a class-wise standard deviation of 5.20%, 0.7% lower than the second lowest, the dual-path DenseNet. This indicates the highest stability of our model's performance among different categories. These findings further affirm the effectiveness of our model and its significant advantage. Thus, these experimental results showcase the superiority and rationale of using ResNet as the backbone.

On the other hand, this paper compared the iteration processes of training models using different networks as backbones. The specific results are depicted in Figure 4, where the horizontal axis represents the number of iterations, and the vertical axis represents accuracy on the validation set. Based on the outcomes in Figure 4, our model outperformed models using other networks as backbones. Although the two-stream DenseNet showed a result similar to the proposed model, its convergence speed was notably slower, which requires approximately 30 iterations to catch up with the proposed model. Notably, our model demonstrated rapid convergence in the early stages of training, which achieves high accuracy with only a few training iterations and maintains a lead. These outcomes underscore the superiority of using ResNet as the backbone concerning training efficiency.

Concurrently, Figure 5 demonstrates a stability test of different networks used as backbones. The boxplot's horizontal axis represents various image classification methods, and the vertical axis represents the accuracy metrics corresponding to each method. The box shape illustrates the data distribution of the corresponding results. While using DenseNet as the backbone achieves the second-highest performance, there is still a noticeable gap compared to the proposed method. In contrast, GoogleNet demonstrates high stability, but its performance notably lags behind the proposed method. Overall, employing ResNet as the backbone maintains high stability while achieving high accuracy. The comparisons in performance, convergence speed, and stability indicate that using ResNet as the backbone allows for faster model training while achieving optimal performance and stability.

4.3 Performance Comparison

In the comparison experiment, each type of Chrysanthemum type dataset is divided into a training set, a validation set and a test set according to the ratio of 6:2:2. To avoid experimental errors due to random sampling, the train-test process was repeated 50 times. The model with the lowest error on the validation set was selected as the final model, and the average values of Acc, Recall, F1, and STD over 50 experiments were

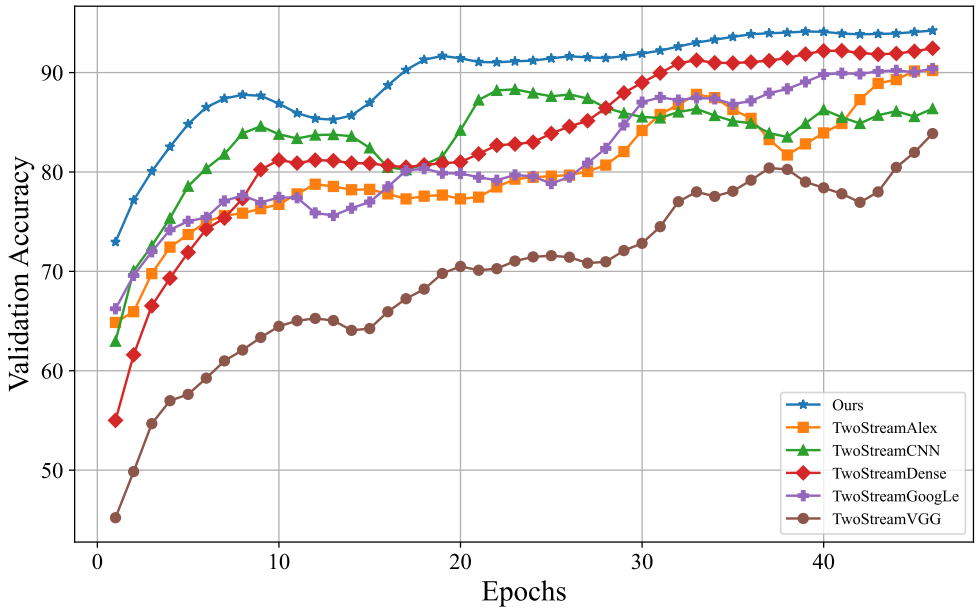


Fig. 4 Convergence speeds of different networks as backbone

calculated and considered as the final results. To demonstrate the effectiveness of the proposed method, this paper compared it against five image classification methods, including BCNN, EfficientNetV2, RegNet, ConvNeXtV2, and TinyVit. Considering the smallest version of ResNet was utilized as the backbone, the corresponding smallest versions of the other models were chosen for comparison, such as EfficientNetV2-s, ConvNeXtV2-atto, etc. Additionally, all models utilized pre-trained parameters from ImageNet to ensure a fair comparison in terms of resource consumption and model complexity.

According to Table 3, the proposed method exhibits competitive performance in accuracy, achieving 93.8%, slightly higher than ConvNeXtV2 and TinyVit's 93.5%, and other classical models such as EfficientNetV2 with 91.7% and RegNet with 92.1%. Moreover, the proposed model displays excellent performance in terms of standard deviation between categories, recording only 5.20, considerably lower compared to other models ranging from 5.61 to 8.52. This not only showcases exceptional overall performance but also highlights the model's stability and balance. These results underscore the effectiveness and robustness of the proposed model, validating its superiority in Chrysanthemum image classification tasks.

Furthermore, this paper recorded the accuracy of each method in every Chrysanthemum category along with the Average Accuracy (AA), as shown in Table 4. From the table, it is obvious that most individual category accuracy of the proposed method in this paper surpasses 90%, with some exceeding 95%, and only 3 categories falling below 90%. The proposed method achieves the highest accuracy in 7 categories, with the highest overall average accuracy of 93.1%. In contrast, ConvNeXtV2's accuracy

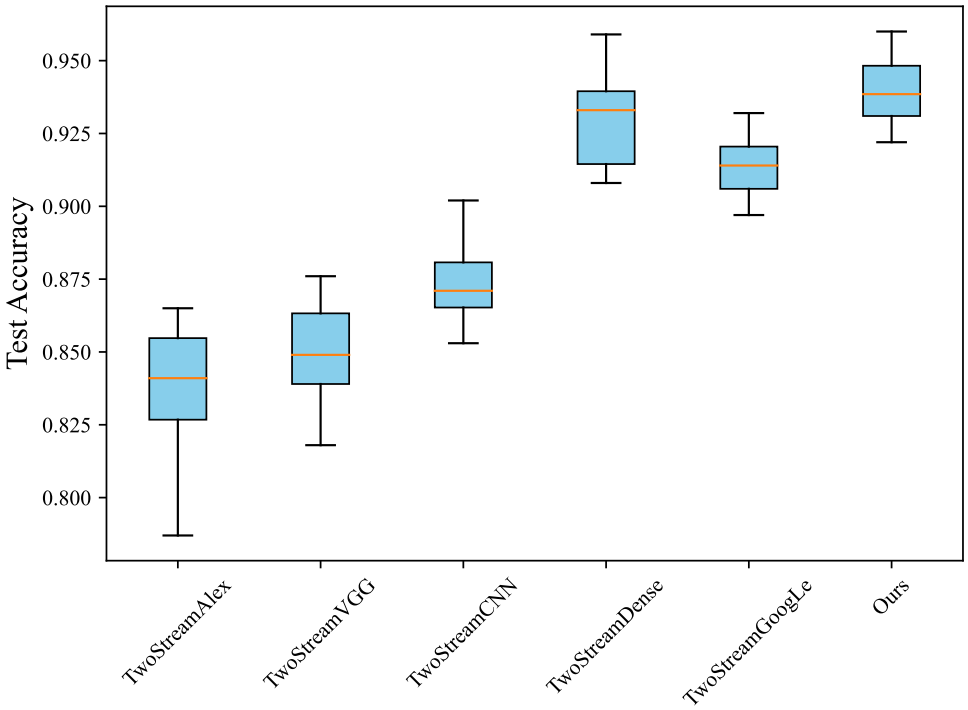


Fig. 5 The accuracy stability of different networks as backbone.

Table 3 Performance comparisons of different image classification methods

Method	Acc(%)	Recall(%)	F1(%)	STD
BCNN(2017)	88.1	88.2	87.4	8.52
RegNet(2020)	92.1	91.7	91.7	6.05
EfficientNetV2(2021)	91.7	91.4	91.3	6.44
ConvNeXtV2(2023)	93.5	92.9	92.9	6.32
TinyVit(2022)	93.5	93.2	93.1	5.61
Ours	93.8	93.5	93.4	5.20

on AHHS-chry5 is below 80% and falls below 90% in 5 categories, while TinyVit has 2 categories with accuracy below 80%. Similarly, EfficientNetV2 achieves the highest accuracy only in YCXY-chry15. This indicates relatively weaker performance of ConvNeXtV2, EfficientNetV2, and TinyVit for certain categories of Chrysanthemum images, while the proposed method exhibits better overall classification ability and more robust performance in specific categories.

On the other hand, this paper compared the iterative processes of different image classification models during training. The specific results are shown in Figure 6, where the horizontal axis represents the number of iterations, and the vertical axis represents

Table 4 Experimental results of different Chrysanthemum categories from different origins.

Category	BCNN	RegNet	EfficientNetV2	TinyVit	ConvNeXtV2	Ours
1	86.0%	84.6%	85.9%	88.5%	85.8%	**91.6%**
2	86.5%	89.8%	91.9%	89.0%	90.2%	**95.1%**
3	75.8%	83.2%	83.2%	**84.2%**	83.9%	81.8%
4	63.5%	74.5%	78.3%	72.9%	75.9%	**82.7%**
5	78.0%	82.4%	83.5%	79.6%	84.9%	**86.7%**
6	99.8%	**100.0%**	98.9%	**100.0%**	**100.0%**	**100.0%**
7	95.9%	96.7%	96.9%	96.7%	**98.0%**	94.3%
8	83.8%	90.2%	91.5%	91.7%	88.3%	**94.2%**
9	98.9%	99.2%	99.4%	99.2%	**100.0%**	99.8%
10	92.9%	94.0%	95.8%	95.4%	**96.0%**	90.3%
11	94.2%	92.8%	96.7%	**97.9%**	95.9%	96.7%
12	93.5%	93.4%	93.2%	90.0%	93.4%	**96.1%**
13	94.3%	98.3%	97.4%	**98.9%**	97.7%	96.1%
14	87.5%	93.8%	94.3%	**95.5%**	92.6%	93.0%
15	87.4%	96.5%	**98.1%**	95.8%	96.5%	97.4%
16	91.2%	94.9%	96.4%	95.2%	**96.4%**	95.8%
17	89.4%	90.4%	91.0%	92.5%	**94.3%**	93.8%
18	77.4%	86.1%	85.4%	**93.5%**	87.0%	90.7%
Avg	87.6%	91.2%	92.1%	92.0%	92.0%	**93.1%**

Table 5 Experimental results of different methods on individual Chrysanthemum categories.

Method	Acc(%)	Recall(%)	F1(%)	STD
BCNN(2017)	92.2	87.3	87.6	10.1
RegNet(2020)	94.7	91.0	91.5	8.16
EfficientNetV2(2021)	95.5	92.3	92.7	6.67
ConvNeXtV2(2023)	94.3	89.8	90.4	8.79
TinyVit(2022)	95.2	91.7	92.4	7.22
Ours	95.8	93.3	93.5	5.79

the accuracy on the validation set. Based on the results in Figure 6, the proposed method demonstrates superior performance compared to other comparative methods. Although EfficientNetV2 and TinyVit converge quickly, their final accuracy falls below that of the proposed model. ConvNeXtV2's curve closely resembles the proposed method 's curve but achieves slightly lower final accuracy. These findings underscore the superior performance and training efficiency of the proposed method.

Additionally, this paper conducted experiments involving the merging of major Chrysanthemum categories and the subsets of Chrysanthemum morifolium 'Hangbaiju'. The experimental results are illustrated in Tables 5 and 6.

As observed from Table 5, the proposed model demonstrates higher performance when dealing with the amalgamation of major Chrysanthemum categories. It achieved 95.8%, 93.3%, and 93.5% in accuracy, recall, and F1 score respectively, surpassing other models' results. Additionally, the proposed model showcases good performance in terms of standard deviation, with a value of only 5.79. These results indicate that the proposed method exhibits better generalization and discrimination capabilities when major Chrysanthemum categories are amalgamated. The observations from Table 6

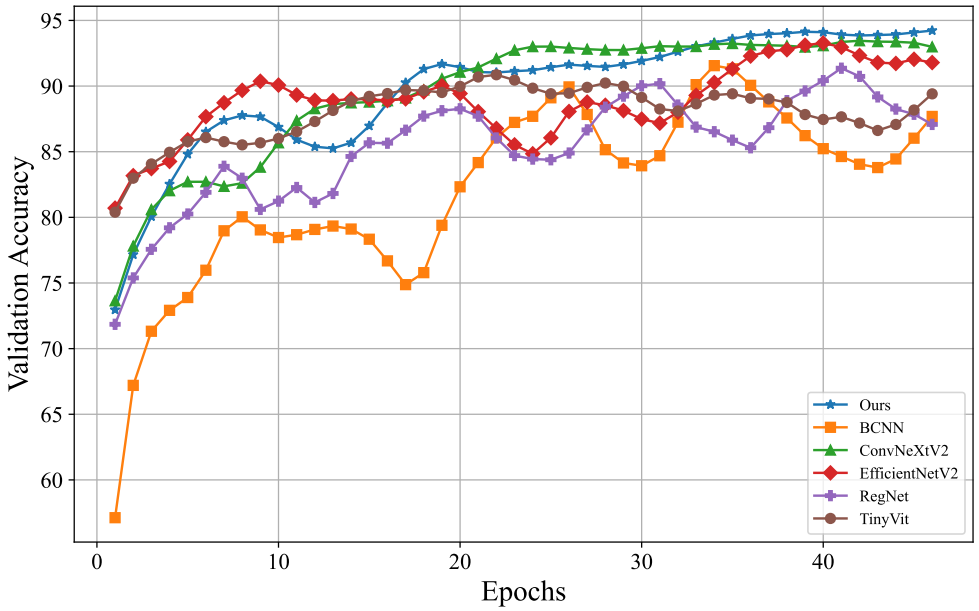


Fig. 6 Convergence speeds of different methods.

Table 6 Experimental results of different methods on Chrysanthemum morifolium 'Hangbaiju' from different origins.

Method	Acc(%)	Recall(%)	F1(%)	STD
BCNN(2017)	90.8	90.5	90.4	4.32
RegNet(2020)	94.1	93.8	93.8	2.75
EfficientNetV2(2021)	95.3	95.0	95.0	2.49
ConvNeXtV2(2023)	93.8	93.1	93.2	4.07
TinyVit(2022)	94.1	93.9	93.7	2.67
Ours	95.9	95.3	95.7	2.60

indicate that within the subcategories of Chrysanthemum morifolium 'Hangbaiju', the proposed model performs exceptionally well, effectively distinguishing between these similar categories. While EfficientNetV2 shows performance metrics close to those of the proposed model, it even outperforms it in terms of standard deviation. However, ConvNeXtV2 and TinyVit do not match the performance of the proposed model.

Analyzing these tables collectively, it's clear that while EfficientNetV2 performs well with fewer categories, its performance diminishes with an increase in categories (as observed in Table 3). Conversely, ConvNeXtV2 and TinyVit show overall accuracy rates similar to the proposed model with more categories, yet exhibit performance drops for specific or fewer categories. The proposed model demonstrates higher stability and reliability in classifying these diverse categories, effectively capturing intra-class feature differences. This superiority isn't just evident in overall accuracy but also

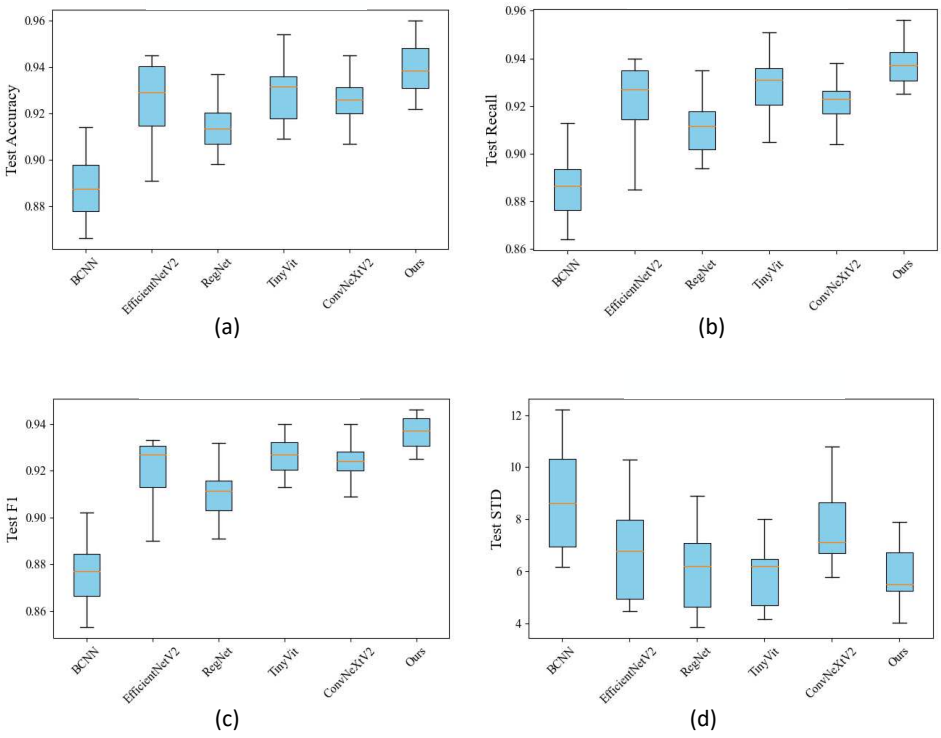


Fig. 7 The stability of different methods.

in discerning and capturing nuanced differences between categories, offering a more dependable and precise solution for Chrysanthemum classification.

4.4 Stability Analysis

To ascertain the stability of our proposed method, this paper created box plots comparing it with other methods in Chrysanthemum classification accuracy. As depicted in Figure 7, the horizontal axis of the box plot represents various image classification methods, and the vertical axis denotes the evaluation metrics corresponding to each method, illustrating the data distribution of the results. Based on the outcomes in (a)(b)(c), EfficientNetV2 and TinyVit attained suboptimal results, albeit with a gap in stability compared to our method outlined in this paper. However, while RegNet and ConvNeXtV2 exhibited higher stability, their performance across various indicators was inferior to the proposed method outlined in this paper. Figure (d) demonstrates that the proposed method excels in stability concerning inter-class variance, while RegNet and TinyVit lag slightly and ConvNeXtV2 diverges significantly from the proposed method. In summary, the proposed method demonstrates both high stability and accuracy, facilitating effective classification of Chrysanthemum images.

Table 7 Experimental results of ablation experiments

Method	Acc(%)	Recall(%)	F1(%)	STD
Front image only	85.5	85.0	85.0	11.7
Back image only	90.5	90.1	90.0	7.61
Both front and back image	91.0	90.0	90.1	8.47
Cross path residual connections	93.0	92.3	92.2	6.25
α=1	91.7	91.2	91.1	8.67
α=1/2	93.3	92.7	92.7	5.57
α=1/4	92.9	92.6	92.8	5.41
α=1/3(ours)	93.8	93.5	93.4	5.20

4.5 Ablation Experiments

To assess the contribution of each component in the Chrysanthemum classification method proposed here, this paper conducted several ablation experiments. These included: 1) using front Chrysanthemum images only, 2) using back Chrysanthemum images only, 3) utilizing both front and back data simultaneously, 4) implementing cross-path residual connections solely in convolutional layers, and 5) adjusting the weight ratio α of the two-stream summation in fully connected layers, with α values of 1, 1/2, 1/3, 1/4 (this paper used α=1/3 in this paper). These experiments aimed to analyze each component's impact on method performance, offering a deeper insight into the method's role and significance in Chrysanthemum classification. Detailed experiment results are outlined in Table 7.

Through observing the results in the table, it can be concluded that solely using front-facing image features results in lower accuracy, reaching only 85.5%. However, employing back-facing image features leads to a significant performance boost, achieving an accuracy of 90.5%. This demonstrates an improvement in performance by using back-facing features alone but doesn't fully exploit the model's potential. Meanwhile, the performance slightly improves to 91.0% by using both front and back image features together, indicating that mere feature concatenation doesn't fully accomplish feature fusion. In contrast, employing cross-path residual connections in the convolutional layers better achieves feature fusion, elevating accuracy to 92.6%. Moreover, the weighted summation operation in the fully connected layers further enhances feature fusion, and we identified the most suitable weight ratio. The 1:3 weighted ratio exhibits the best performance, reaching the highest accuracy of 93.8%. These findings highlight the pivotal role of feature fusion in Chrysanthemum classification methods, offering significant insights into method performance. Overall, the comprehensive use of cross-path residual connections and fully connected weighted summation effectively fuse image features from both sides of Chrysanthemum images. This approach accounts for the differences between front and back features among different Chrysanthemum types, achieving feature complementarity for more accurate Chrysanthemum classification.

4.6 Visual Analysis

The paper utilizes t-SNE to generate a scatter plot illustrating the clustering of Chrysanthemum classifications, as shown in Figure 8. The x-axis and y-axis represent

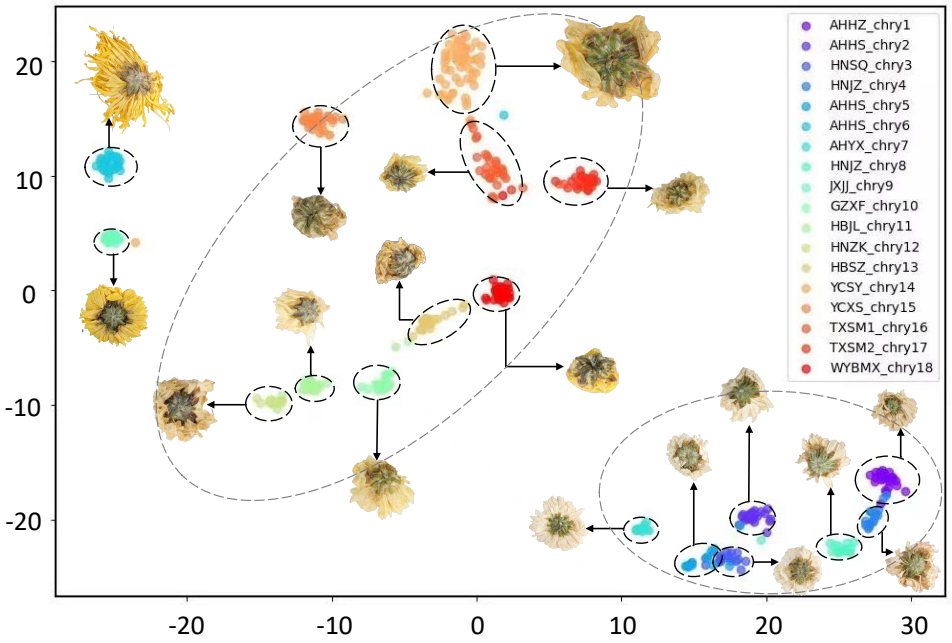


Fig. 8 Visual analysis for Chrysanthemum classification.

specific features, with each point representing an individual Chrysanthemum sample. Different colored points denote distinct Chrysanthemum categories, displaying their distribution in the feature space. The visualization reveals two primary clusters within the Chrysanthemum dataset, accompanied by smaller clusters inside them, depicting similar shapes and corresponding to different Chrysanthemum categories. The first prominent cluster includes Hangbaiju like GZXF_chry10, HBJL_chry11, HBSZ_chry13, YCSY_chry14. The second significant cluster includes Chrysanthemums like AHBZ_chry1, AHHS_chry2, HNSQ_chry3, HNJZ_chry4. This visualization offers valuable insights into studying the differences among Chrysanthemum categories, facilitating model optimization for better category differentiation.

5 Conclusion

This paper proposes a Chrysanthemum classification method based on the fusion of deep visual features of both the front and back sides. Different Chrysanthemum images are collected and labeled with origins and classifications. The front and back images underwent separate preprocessing and were used as inputs for a two-stream network. Leveraging single-path residual connections and cross-path residual connections expands the receptive field of the network and fully fuses the features of the

front and back sides. Overall, the innovation of this paper lies in the introduced two-stream deep neural network model and the strategy of fusing deep features of the front and back sides of Chrysanthemums. The proposed method demonstrates higher accuracy and stronger stability, offering an effective and reliable solution for classifying the origins of Chrysanthemums. Despite some shortcomings in our model's performance on certain categories, we plan to explore the effects of texture feature extraction and enhance data augmentation strategies in future research to further improve classification performance.

Declarations

- Ethics approval and consent to participate
  Not applicable
- Consent for publication
  Not applicable
- Competing interests
  The authors have declared that there are no competing interests exist.
- Ethical standard
  The authors have no conflict of interest to disclose. Further, the authors certify that, the research presented in this article does not involve any human participants or animals. Moreover, all the datasets used in our experiment are public.
- Availability of data and materials
  Data supporting the results of this study are available from the corresponding author upon reasonable request, including source code and experimental results.
- Authors' contributions
  Conceived and designed the experiments: Xichen Yang and Yifan Chen. Performed the experiments: Xichen Yang and Yifan Chen. Analyzed the data: Hui Yan, Yifan Chen, Jia Liu, Jian Jiang, Zhongyuan Mao. Wrote and reviewed the paper: Xichen yang, Hui Yan, Yifan Chen, Jian Jiang, Zhongyuan Mao, Tianshu Wang, Jia Liu. Approved the final version of the paper: Xichen yang, Hui Yan, Yifan Chen, Tianshu Wang.
- Acknowledgements
  The authors are grateful for the financial support provided by National Natural Science Foundation of China (Grant No. 62101268), National Natural Science Foundation of China (Grant No. 82204770), Youth Science Foundation of Jiangsu Province (Grant No. BK20210696), High Level Key Discipline Construction Project of the National Administration of Traditional Chinese Medicine - Resource Chemistry of Chinese Medicinal Materials (No. zyyzdxk-2023083), Innovation Team and Talents Cultivation Program of National Administration of Traditional Chinese Medicine (ZYYCXTD-D-202005), China Agriculture Research System of MOF and MARA(Grant No. CARS-21) , Jiangsu Province 333 High-level Talents Training Project, and "Qing Lan Project" in Colleges and universities in Jiangsu.

References

[1] Sun, J., Wang, Z., Chen, L., Sun, G.: Hypolipidemic effects and preliminary mechanism of chrysanthemum flavonoids, its main components luteolin and luteoloside in hyperlipidemia rats. Antioxidants **10**(8), 1309 (2021)

[2] Zhan, J., He, F., Cai, H., Wu, M., Xiao, Y., Xiang, F., Yang, Y., Ye, C., Wang, S., Li, S.: Composition and antifungal mechanism of essential oil from chrysanthemum morifolium cv. fubaiju. Journal of Functional Foods **87**, 104746 (2021)

[3] Chen, S., Liu, J., Dong, G., Zhang, X., Liu, Y., Sun, W., Liu, A.: Flavonoids and caffeoylquinic acids in chrysanthemum morifolium ramat flowers: A potentially rich source of bioactive compounds. Food chemistry **344**, 128733 (2021)

[4] Jiang, S., Wang, M., Jiang, Z., Zafar, S., Xie, Q., Yang, Y., Liu, Y., Yuan, H., Jian, Y., Wang, W.: Chemistry and pharmacological activity of sesquiterpenoids from the chrysanthemum genus. Molecules **26**(10), 3038 (2021)

[5] He, J., Zhu, S., Chu, B., Bai, X., Xiao, Q., Zhang, C., Gong, J.: Nondestructive determination and visualization of quality attributes in fresh and dry chrysanthemum morifolium using near-infrared hyperspectral imaging. Applied Sciences **9**(9), 1959 (2019)

[6] Youssef, F.S., Eid, S.Y., Alshammari, E., Ashour, M.L., Wink, M., El-Readi, M.Z.: Chrysanthemum indicum and chrysanthemum morifolium: chemical composition of their essential oils and their potential use as natural preservatives with antimicrobial and antioxidant activities. Foods **9**(10), 1460 (2020)

[7] Yang, H.-M., Sun, C.-Y., Liang, J.-L., Xu, L.-Q., Zhang, Z.-B., Luo, D.-D., Chen, H.-B., Huang, Y.-Z., Wang, Q., Lee, D.Y.-W., *et al.*: Supercritical-carbon dioxide fluid extract from chrysanthemum indicum enhances anti-tumor effect and reduces toxicity of bleomycin in tumor-bearing mice. International journal of molecular sciences **18**(3), 465 (2017)

[8] Yuan, H., Jiang, S., Liu, Y., Daniyal, M., Jian, Y., Peng, C., Shen, J., Liu, S., Wang, W.: The flower head of chrysanthemum morifolium ramat.(juhua): A paradigm of flowers serving as chinese dietary herbal medicine. Journal of Ethnopharmacology **261**, 113043 (2020)

[9] Luo, D., Chen, J., Gao, L., Liu, Y., Wu, J.: Geographical origin identification and quality control of chinese chrysanthemum flower teas using gas chromatography–mass spectrometry and olfactometry and electronic nose combined with principal component analysis. International Journal of Food Science & Technology **52**(3), 714–723 (2017)

[10] Liu, H.-x., Zhou, Q., Sun, S.-q., Bao, H.-j.: Discrimination of different chrysanthemums with fourier transform infrared spectroscopy. Journal of Molecular Structure **883**, 38–47 (2008)

[11] Long, Z., Ai-Li, Q., Han, J., Li, Z., Qiong, J., Jia-Rong, P., Cheng, Z., University, L.: Discriminant the geographical origin of hangzhou white chrysanthemum based on mineral elements. Journal of Nuclear Agricultural Sciences **27**(10), 1553–1559 (2013)

[12] Chen, Y., Zhen, X.-T., Yu, Y.-L., Shi, M.-Z., Cao, J., Zheng, H., Ye, L.-H.: Chemoinformatics based comprehensive two-dimensional liquid chromatography-quadrupole time-of-flight mass spectrometry approach to chemically distinguish chrysanthemum species. Microchemical Journal **168**, 106464 (2021)

[13] Platt, J.: Sequential minimal optimization: A fast algorithm for training support vector machines (1998)

[14] Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., Fei-Fei, L.: Imagenet: A large-scale hierarchical image database. In: 2009 IEEE Conference on Computer Vision and Pattern Recognition, pp. 248–255 (2009). Ieee

[15] Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. Advances in neural information processing systems **25** (2012)

[16] Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556 (2014)

[17] Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A.: Going deeper with convolutions. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1–9 (2015)

[18] He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778 (2016)

[19] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I.: Attention is all you need. Advances in neural information processing systems **30** (2017)

[20] Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., et al.: An image is worth 16x16 words: Transformers for image recognition at scale. arXiv preprint arXiv:2010.11929 (2020)

[21] Momeny, M., Neshat, A.A., Jahanbakhshi, A., Mahmoudi, M., Ampatzidis, Y.,

Radeva, P.: Grading and fraud detection of saffron via learning-to-augment incorporated inception-v4 cnn. Food Control **147**, 109554 (2023)

[22] Hadipour-Rokni, R., Asli-Ardeh, E.A., Jahanbakhshi, A., Sabzi, S., *et al.*: Intelligent detection of citrus fruit pests using machine vision system and convolutional neural network through transfer learning technique. Computers in Biology and Medicine **155**, 106611 (2023)

[23] Momeny, M., Jahanbakhshi, A., Neshat, A.A., Hadipour-Rokni, R., Zhang, Y.-D., Ampatzidis, Y.: Detection of citrus black spot disease and ripeness level in orange fruit using learning-to-augment incorporated deep networks. Ecological Informatics **71**, 101829 (2022)

[24] Azadnia, R., Fouladi, S., Jahanbakhshi, A.: Intelligent detection and waste control of hawthorn fruit based on ripening level using machine vision system and deep learning techniques. Results in Engineering **17**, 100891 (2023)

[25] Zhai, G., Li, Z., Lu, W., Zhao, Y., Wang, C.: Study on varieties of identification ornamental chrysanthemum based on image processing. Journal of Chinese Agricultural Mechanization **37**(2), 105–110 (2016)

[26] Liu, Z., Gao, K., Tian, y., Dai, S., Song, X.: Identification of chrysanthemum cultivars based on unfolding image with lbp texture feature. Research progress of ornamental horticulture in China (2017)

[27] Lin, T.-Y., RoyChowdhury, A., Maji, S.: Bilinear cnn models for fine-grained visual recognition. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 1449–1457 (2015)

[28] Radosavovic, I., Kosaraju, R.P., Girshick, R., He, K., Dollár, P.: Designing network design spaces. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 10428–10436 (2020)

[29] Tan, M., Le, Q.: Efficientnet: Rethinking model scaling for convolutional neural networks. In: International Conference on Machine Learning, pp. 6105–6114 (2019). PMLR

[30] Hu, J., Shen, L., Sun, G.: Squeeze-and-excitation networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 7132–7141 (2018)

[31] Woo, S., Park, J., Lee, J.-Y., Kweon, I.S.: Cbam: Convolutional block attention module. In: Proceedings of the European Conference on Computer Vision (ECCV), pp. 3–19 (2018)

[32] Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S., Guo, B.: Swin transformer: Hierarchical vision transformer using shifted windows. In: Proceedings of

the IEEE/CVF International Conference on Computer Vision, pp. 10012–10022 (2021)

[33] Wu, K., Zhang, J., Peng, H., Liu, M., Xiao, B., Fu, J., Yuan, L.: Tinyvit: Fast pretraining distillation for small vision transformers. In: European Conference on Computer Vision, pp. 68–85 (2022). Springer

[34] Woo, S., Debnath, S., Hu, R., Chen, X., Liu, Z., Kweon, I.S., Xie, S.: Convnext v2: Co-designing and scaling convnets with masked autoencoders. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 16133–16142 (2023)

[35] Liu, Z., Mao, H., Wu, C.-Y., Feichtenhofer, C., Darrell, T., Xie, S.: A convnet for the 2020s. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 11976–11986 (2022)

[36] He, K., Chen, X., Xie, S., Li, Y., Dollár, P., Girshick, R.: Masked autoencoders are scalable vision learners. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 16000–16009 (2022)

[37] Long, W., Wang, S.-R., Suo, Y., Chen, H., Bai, X., Yang, X., Zhou, Y.-P., Yang, J., Fu, H., *et al.*: Fast and non-destructive discriminating the geographical origin of hangbaiju by hyperspectral imaging combined with chemometrics. Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy **284**, 121786 (2023)

[38] Yuan, P., Ren, S., Zhai, Z., Xu, H.: Chrysanthemum phenotypic classification based on semi-supervised active learning. Nongye Jixie Xuebao/Transactions of the Chinese Society of Agricultural Machinery **49**(9) (2018)

[39] Yuan, P., Qian, S., Zhai, Z., FernánMartínez, J., Xu, H.: Study of chrysanthemum image phenotype on-line classification based on transfer learning and bilinear convolutional neural network. Computers and Electronics in Agriculture **194**, 106679 (2022)

[40] Liu, Z., Wang, J., Tian, Y., Dai, S.: Deep learning for image-based large-flowered chrysanthemum cultivar recognition. Plant Methods **15**(1), 1–11 (2019)

[41] Wang, J., Tian, Y., Zhang, R., Liu, Z., Tian, Y., Dai, S.: Multi-information model for large-flowered chrysanthemum cultivar recognition and classification. Frontiers in Plant Science **13**, 806711 (2022)

[42] Huang, S., Liu, G.: Research on fine-grained classification of chrysanthemum images based on multi-scale and multi-parallel convolutional neural network. In: International Conference on Cyber Security, Artificial Intelligence, and Digital Economy (CSAIDE 2023), vol. 12718, pp. 430–437 (2023). SPIE

[43] Ma, Z., Oh, C.: A wavelet-based dual-stream network for underwater image enhancement. In: ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 2769–2773 (2022). IEEE

[44] Pan, Z., Yu, W., Lei, J., Ling, N., Kwong, S.: Tsan: Synthesized view quality enhancement via two-stream attention network for 3d-hevc. IEEE Transactions on Circuits and Systems for Video Technology **32**(1), 345–358 (2021)

[45] Liu, W., Qian, J., Yao, Z., Jiao, X., Pan, J.: Convolutional two-stream network using multi-facial feature fusion for driver fatigue detection. Future Internet **11**(5), 115 (2019)

[46] Zhou, A., Ma, Y., Ji, W., Zong, M., Yang, P., Wu, M., Liu, M.: Multi-head attention-based two-stream efficientnet for action recognition. Multimedia Systems **29**(2), 487–498 (2023)

[47] Wang, Q., Huang, W., Xiong, Z., Li, X.: Looking closer at the scene: Multiscale representation learning for remote sensing image scene classification. IEEE Transactions on Neural Networks and Learning Systems **33**(4), 1414–1428 (2020)

[48] Tu, Z., Zhang, J., Li, H., Chen, Y., Yuan, J.: Joint-bone fusion graph convolutional network for semi-supervised skeleton action recognition. IEEE Transactions on Multimedia (2022)

[49] Zheng, J., Liu, H., Feng, Y., Xu, J., Zhao, L.: Casf-net: Cross-attention and cross-scale fusion network for medical image segmentation. Computer Methods and Programs in Biomedicine **229**, 107307 (2023)

[50] Xie, X., Zhang, W., Pan, X., Xie, L., Shao, F., Zhao, W., An, J.: Canet: Context aware network with dual-stream pyramid for medical image segmentation. Biomedical Signal Processing and Control **81**, 104437 (2023)

[51] Ma, H., Cui, Z., Gan, Z., Tang, G., Liu, F.: Saliency-enhanced two-stream convolutional network for no-reference image quality assessment. Journal of Electronic Imaging **31**(2), 023010–023010 (2022)

[52] Chen, F., Fu, H., Yu, H., Chu, Y.: Using hvs dual-pathway and contrast sensitivity to blindly assess image quality. Sensors **23**(10), 4974 (2023)