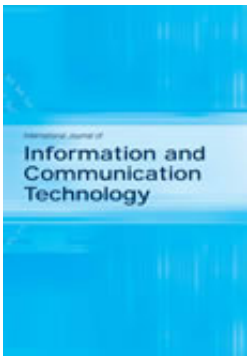


International Journal of Information and Communication Technology

ISSN online: 1741-8070 - ISSN print: 1466-6642
https://www.inderscience.com/ijict

Ink painting classification method based on deep feature fusion

Yiwen Chen

DOI: 10.1504/IJICT.2025.10072363

Article History:

Received:	24 May 2025
Last revised:	12 June 2025
Accepted:	13 June 2025
Published online:	25 July 2025

Ink painting classification method based on deep feature fusion

Yiwen Chen

School of Architectural Decoration and Art Design,
Jiangsu Vocational College of Electronics and Information,
Huai'an 223003, China
Email: cyw20200313@163.com

Abstract: In response to the problem of insufficient classification accuracy caused by the neglect of multi-scale feature fusion in traditional convolutional neural networks (CNN) for ink painting classification tasks, this paper proposes a deep feature fusion based ink painting classification model. Firstly, a multi-scale feature extraction module is constructed to capture the stroke details and composition features of ink paintings through parallel convolutional kernels with different receptive fields. Secondly, a dual path attention fusion module is designed, which adopts a parallel mechanism of channel attention and spatial attention to achieve adaptive weighted fusion of cross level features, enhancing the feature expression ability of ink wash blending effect; simultaneously introducing a cross layer connection structure to promote the interaction and fusion of shallow texture features and deep semantic features. This study provides new technological ideas for intelligent appreciation in the protection of digital cultural heritage.

Keywords: deep feature fusion; classification of ink painting; attention fusion.

Reference to this paper should be made as follows: Chen, Y. (2025) 'Ink painting classification method based on deep feature fusion', *Int. J. Information and Communication Technology*, Vol. 26, No. 28, pp.103–117.

Biographical notes: Yiwen Chen received her Master degree from Nanjing Normal University in June 2018. She is currently a teacher in the Jiangsu Vocational College of Electronics and Information. Her research interests include arts design, ink painting, feature extraction and convolutional neural networks.

1 Introduction

As a treasure of traditional Chinese art, ink painting carries a thousand years of cultural accumulation and aesthetic philosophy (Zeng et al., 2024). Its unique brush and ink techniques, as well as the artistic expression of the interplay between reality and virtuality, have formed a visual feature that is completely different from Western oil painting (Fan et al., 2023). With the increasingly urgent need for the protection of digital cultural heritage, how to use artificial intelligence technology to automate the analysis and classification of ink paintings has become an important issue in the intersection of computer vision and digital art. However, the artistic characteristics of ink painting, such as the gradient of ink colour blending and the abstraction of blank composition, pose

unique challenges to its classification task: the successful experience of traditional convolutional neural networks in natural image classification is difficult to directly transfer, especially the problem of insufficient multi-scale feature fusion, which limits the model's ability to capture key artistic elements of ink painting.

In recent years, deep learning techniques have made significant progress in the field of art classification. The classic model proposed by He et al. (2015) extracts high-order semantic features through stacked convolutional layers and performs well in art style classification tasks such as oil painting and printmaking. However, the particularity of ink painting poses three challenges to existing methods:

1 Multi scale feature dependency: the stroke details of ink painting require local fine-grained features, while the composition mood relies on global contextual perception, and a single scale convolution kernel is difficult to balance

2 Dynamic feature interaction requirements: the colour gradient formed by ink blending has unstructured features, requiring the model to dynamically adjust the fusion weights of different levels of features

3 Low contrast sensitivity: traditional CNNs are sensitive to high contrast edges, but ink painting uses ink colour intensity as the core expression method, and the cross layer interaction between shallow textures and deep semantic features is crucial.

Current research mainly focuses on data augmentation or deep learning strategies, but there is insufficient attention paid to optimising feature fusion mechanisms. Sandoval et al. (2019) introduced a new two-stage image classification method aimed at improving the accuracy of style classification; Chen (2020) designed and improved an ink painting rendering algorithm based on deep learning framework and convolutional neural network model. But there is still an urgent need to design a specialised deep feature fusion framework for the characteristics of ink painting.

In recent years, feature fusion technology has been widely applied in fields such as medical image segmentation and remote sensing image interpretation (Ma et al., 2023). Typical methods include: multi-scale feature pyramid: fusing different resolution features through top-down paths, but the concatenated structure may result in the loss of shallow detail information; Attention guided fusion: using channel or spatial attention weighted feature maps, but a single attention branch is difficult to capture the complex coupling relationship between texture and semantics in ink paintings; Dense connection architecture: enhances feature reuse through cross layer connections, but treating all hierarchical features equally may introduce redundant noise.

Although these methods have improved the feature expression ability in general scenarios, there are still significant shortcomings when directly applied to ink painting classification. Firstly, the traditional pyramid structure assumes that feature scales are abstracted layer by layer from shallow to deep, and the stroke details of ink painting may degrade in deep networks; secondly, existing attention mechanisms often use serial computation (channel first, spatial later, or vice versa), which cannot synchronously model the collaborative characteristics of ink colour distribution (channel dimension) and composition layout (spatial dimension) in ink painting.

In response to the above issues, this article proposes a deep feature fusion driven ink painting classification model, with core innovations including:

1 Multi scale heterogeneous convolution module: by using parallel dilated convolution and conventional convolution, it captures local details of strokes (small receptive field) and global correlations of composition (large receptive field) at the same computational cost

2 Dual path attention fusion mechanism: design parallel channel space attention subnets to separately model ink density distribution and stroke space topology, and achieve adaptive fusion of cross level features through dynamic gating units

3 Cross layer gradient enhanced connection: embedding learnable weight coefficients in skip connections to suppress shallow noise while enhancing cross layer propagation of key texture features.

2 Analysis of characteristics of ink painting

As a unique art form in China, ink painting has distinct cultural characteristics in its creative medium and expressive techniques. From the perspective of material carriers, ink painting is based on rice paper and silk, and with the help of the elasticity of the brush and the permeability of ink colour, it produces ink level changes of 'focal, thick, heavy, light, and clear' through the proportion control of water and ink. This material characteristic determines that the visual expression of ink painting is different from the overlay style of Western oil painting, but is dominated by the diffusion, blending, and infiltration of water as a medium (Hu, 2023). In terms of brushstrokes, artists use techniques such as centre, side, and reverse strokes to form basic brushstrokes such as 'texturing, rubbing, pointing, and dyeing'. Among them, texturing is the core technique for expressing the texture of mountains and rocks, giving rise to dozens of variations such as linen texturing, axe splitting texturing, and cloud rolling texturing. The texture direction and density rhythm not only follow the physical laws of natural objects, but also carry the subjective interests of the creator. The diversity of this technique makes the local textures of ink painting highly unstructured, and works of the same category may present completely different microstructures due to differences in the author's brushwork habits (Wu et al., 2023).

The construction of ink colour hierarchy in ink painting is the core dimension of its artistic expression. The variation of ink colour intensity is achieved through the control of moisture, forming a gradient and transitional blending effect between rice paper fibres. This 'ink divided into five colours' grayscale expression breaks through the flatness of a single colour and creates a three-dimensional depth of virtual and real space in a two-dimensional image. For example, when depicting mountain scenery shrouded in clouds and mist, artists often use the technique of 'splashing ink' to render large areas, using the edge blurring effect formed by the natural flow of ink to express air perspective; when depicting close-up rocks, the 'accumulation of ink' method is used to layer by layer to enhance the thickness of the texture. This dynamic ink blending process results in non-uniform gradient features in local areas of the image. Traditional computer vision algorithms based on edge detection or region segmentation often struggle to accurately define the boundaries of the body, and overly deep convolutional neural networks may lose subtle ink gradient information due to multiple downsampling operations (Dong and Dechsubha, 2024).

At the composition level, ink painting follows the aesthetic principle of 'using white as black'. Leaving white space is not only the absence of elements in the picture, but also a carrier for extending the artistic conception. The 'corner scenery' of Yuan Ma and Gui Xia in the Southern Song Dynasty often implied the vastness of rivers, lakes, and seas with a large number of blank spaces (Yan et al., 2022). This spatial layout of virtual and real coexistence requires classification models to have the ability to analyse asymmetric compositions and abstract symbols. At the same time, the objects in ink paintings are often highly refined and transformed, such as the birds depicted by the Eight Great Mountain People with minimalist lines to outline their expressions. The abstract and exaggerated nature of their shapes renders traditional shape template matching methods ineffective. In addition, the embedding of textual elements such as captions and seals further increases the complexity of image semantics. These textual areas are not only components of the composition of the image, but also carry independent literary value. However, existing visual models often consider them as interference noise rather than classification criteria.

Traditional convolutional neural networks face multiple adaptive barriers when processing the aforementioned features. Firstly, the fixed receptive field of standard convolution kernels is difficult to accommodate the coexistence of macro imagery and micro brushstrokes in ink painting: smaller convolution kernels (such as 3 × 3) can capture the local texture of the texturing method, but cannot perceive the global spatial relationship formed by the blank composition; expanding the receptive field can improve contextual modelling, but it can lead to a loss of smoothness in stroke details. Secondly, the hierarchical feature extraction mechanism of mainstream network architecture fundamentally conflicts with the feature distribution of ink painting: the edge and texture features extracted by shallow networks have weak response to ink colour blending areas, while the advanced semantic features abstracted by deep networks are difficult to distinguish the brushwork styles of different artists. More importantly, traditional models often use simple concatenation or addition operations for feature fusion, failing to establish dynamic correlations between heterogeneous features such as brushstrokes, ink colour, and composition. For example, although Wei Xu's cursive brushstrokes and Baishi Qi's meticulous brushstrokes belong to the category of freehand brushwork, their feature combination patterns have significant differences, requiring the model to adaptively adjust the interaction weights of different feature channels. These limitations reveal the essential flaws of existing methods in the analysis of ink painting – mechanically transplanting the feature extraction paradigm of Western oil painting into the context of Eastern art, ignoring the unique forms of expression and evaluation system of traditional Chinese aesthetics.

3 The limitations of traditional CNN

The unique artistic language of ink painting poses a fundamental challenge to the architecture design of traditional convolutional neural networks. Although CNN demonstrates strong feature extraction capabilities in natural image classification, its underlying assumptions are significantly misaligned with the visual characteristics of ink painting (Bhatt et al., 2021). Classic models such as ResNet and VGG use stacked convolutional layers to construct hierarchical feature abstractions. This design is effective in processing concrete, high contrast natural images, but its limitations gradually become

exposed when dealing with features such as ink colour gradients and unstructured brushstrokes in ink painting. For example, the fixed size receptive field of standard convolution kernels is difficult to adapt to the characteristics of multi-scale features coexisting in ink painting: although the 3 × 3 small convolution kernel can capture the local texture details of the 'texturing' strokes, it cannot perceive the macro spatial relationships formed by the 'blank space' composition; If a 5 × 5 or 7 × 7 large convolution kernel is used to expand the receptive field, although it can improve the global context modelling ability, it will lead to blurred stroke edges, especially when expressing the 'flying white' technique, where subtle changes in ink colour are easily smoothed. The contradictory nature of this scale selection stems from the multimodal characteristics of ink painting art. The traditional CNN's single path feature extraction mechanism lacks the ability to dynamically adjust the receptive field, making it difficult to achieve a balance between preserving local details and perceiving global context (Kattenborn et al., 2021).

The increase in network depth further exacerbates the distortion problem of feature expression. The mainstream CNN uses pooling layers and stride convolution to achieve spatial downsampling, which was originally designed to extract translation invariant high-order semantic features. However, the ink colour hierarchy of ink paintings will experience irreversible information loss during multiple downsampling processes (Salehi et al., 2023). Taking the 'splashing ink' technique as an example, the gradient area formed by the diffusion of ink colour contains rich grayscale transition information. These features can still maintain high resolution in shallow networks, but as the network depth increases, continuous spatial details are compressed into abstract feature vectors, leading to a gradual degradation of the hierarchical sense of ink colour blending. More seriously, deep networks tend to learn discriminative features strongly related to categories, while the artistic value of ink painting is often reflected in intermediate level features such as brushstrokes and ink techniques.

The simplification of feature fusion strategy is another key bottleneck. Existing models often use concatenation or element wise addition to achieve cross layer feature fusion. These operations assume that features at different levels have spatial alignment and channel homogeneity, but the strokes, ink colours, composition, and other features of ink painting are essentially heterogeneous modalities (Wang et al., 2023). Taking Wei Xu's freehand flowers in the Ming dynasty and Baishi Qi's meticulous brushwork of grass and insects as examples, although both belong to the category of flower and bird painting, the former relies on the dynamic characteristics of wild brushstrokes, while the latter emphasises the structural features of fine contours. Traditional fusion methods treat all feature channels equally and cannot dynamically adjust fusion weights based on input sample characteristics, resulting in insufficient adaptability of the model to differentiated artistic expression techniques. In addition, there is a semantic correlation between the text areas such as inscriptions and seals in ink paintings and the painting subject, but conventional CNNs lack cross modal feature interaction mechanisms, often treating the text areas as interference noise rather than auxiliary classification clues, resulting in insufficient information utilisation (Purwono et al., 2022).

The weak responsiveness to low contrast features exposes the limitations of traditional activation functions. Ink painting is characterised by the use of varying shades of ink, with important features often found in low contrast areas such as distant mountain contours rendered in light ink or transitional zones where ink blends (Dong and Dechsubha, 2024). However, commonly used activation functions such as ReLU have a

hard truncation effect on feature responses close to zero, resulting in the suppression of weak but discriminative ink colour change signals. At the same time, traditional CNN channel attention mechanisms (such as SE modules) are typically based on global average pooling to generate weights, which blurs the spatial specificity of local ink colour distribution in ink paintings through global compression operations. For example, when depicting a landscape scene shrouded in morning mist, key discriminative features may be concentrated in the ink gradient zone in the centre of the image, while low information features in the surrounding blank areas may receive unreasonable weight allocation in the channel weighting process. This bias will reduce the model's ability to focus on core artistic elements.

The interpretability deficiency of the model further limits the application value of traditional CNN in ink painting analysis. Although techniques such as class activation mapping can visualise the attention regions of models, the heat maps generated by existing methods often deviate from the key features that art critics focus on (Marchesi et al., 2021). For example, when identifying the authenticity of works by the Eight Great Mountain People, experts usually judge based on the stroke strength of bird pupils and the degree of penetration of feather ink colour, while CNN models may mistakenly classify based on captions or irrelevant background textures at the edges of the image. This disconnect in semantic understanding reflects the fundamental gap between the traditional CNN feature learning mechanism and the aesthetic evaluation system of ink painting. When the model cannot establish an interpretable correlation between low-level stroke features and high-level artistic value, its classification results are difficult to be truly accepted as a reliable analytical tool in the field of cultural heritage protection (Zhong and Huang, 2022).

These limitations collectively reveal the fundamental contradiction of traditional CNN in the classification of ink painting: its feature extraction paradigm designed on the implicit premise of Western realism painting deeply conflicts with the aesthetic concept of ‘emphasising meaning over form’ in Chinese ink painting. From the geometric constraints of convolutional kernels to the static strategy of feature fusion, from the information loss of downsampling to the spatial insensitivity of attention mechanisms, every design step is deconstructing the unique artistic integrity of ink painting. The mismatch between this technological paradigm and art ontology not only leads to a bottleneck in classification performance, but also hinders the deep dialogue between artificial intelligence and traditional Chinese art philosophy.

4 Deep feature fusion

The core idea of deep feature fusion originates from the hierarchical information integration mechanism of the biological visual system. When processing complex visual scenes, the human brain does not rely on a single level of neurons to complete feature analysis, but relies on multi-level linkage from the V1 area to the IT area in the visual cortex, gradually integrating local edge information into global semantic representations. The primary visual cortex works together with simple and complex cells that are sensitive to direction, capturing detailed features such as line direction and constructing the overall outline of objects through hierarchical transmission. This biological mechanism suggests that artificial neural networks need to break the isolation of interlayer features and establish cross level feature dialogue channels. In the analysis of ink painting, the

micro texture of brushstrokes and the macro artistic conception of composition form a multi-level artistic expression. The one-way feature abstraction process of traditional convolutional neural networks is prone to premature loss of shallow details, while deep feature fusion introduces cross layer connections, allowing deep networks to access high-resolution features of shallow layers in reverse, forming a feedback adjustment mechanism similar to the visual cortex, thus achieving multi granularity analysis of ink language.

From the perspective of information theory, the essence of feature fusion is to optimise and break through information bottlenecks. Shannon entropy theory states that the process of information transmission inevitably involves the screening of useful information and the removal of noise. Traditional neural networks compress spatial dimensions through pooling operations, which can improve the translation invariance of features, but may discard key light ink gradient information in ink paintings along with background noise. The attention mechanism introduced by deep feature fusion can be regarded as a dynamic information gating system, which achieves information entropy redistribution by calculating the importance weights of feature channels and spatial positions. For example, when dealing with the 'splashing ink' technique, the system will automatically enhance the weight of the ink diffusion edge area and suppress the interference signal of the rice paper background. This adaptive selection mechanism makes the information compression process no longer a simple geometric dimensionality reduction, but an intelligent information purification based on task requirements, effectively alleviating the limitations of the 'one size fits all' feature processing in traditional methods.

The theoretical basis of multi-scale feature fusion can be traced back to the wavelet analysis concept in the field of signal processing. Wavelet transform decomposes signals into multi-resolution through basis functions of different scales, capturing high-frequency details while preserving the characteristics of low-frequency contours. It cleverly corresponds to the short-range stroke interlacing and long-range spatial resonance of the 'texturing method' in ink painting. The dilated convolution and multi branch structure in deep neural networks essentially construct an artificial scale space, simulating the multi-scale analysis capability of wavelet transform through parallel convolution kernels with different receptive fields. When dealing with 'lotus leaf patterns' with self similar structures, small-scale convolution focuses on local patterns of short line interlacing, while large-scale convolution captures the overall pattern of texture direction. The collaborative work of the two overcomes the field of view limitations of a single convolution kernel. This multi-scale representation ability is particularly important for analysing the fractal structure commonly found in ink paintings. Through hierarchical feature recombination, it achieves continuous expression from micro brushstrokes to macro artistic conception.

The theoretical support for attention mechanism comes from the visual attention model in cognitive science. When humans appreciate ink painting, they naturally switch the visual focus between elements such as stroke intensity, ink colour hierarchy, and composition centre of gravity. The dual path attention mechanism simulates this cognitive process, with the channel attention path focusing on the rhythmic changes in ink colour like an art critic, and the spatial attention path interpreting the positional relationships of objects like a composition analyst. The parallel operation of the two breaks through the sequential limitations of traditional attention mechanisms, enabling the model to synchronously process the coupled ink rhyme and spatial features in ink

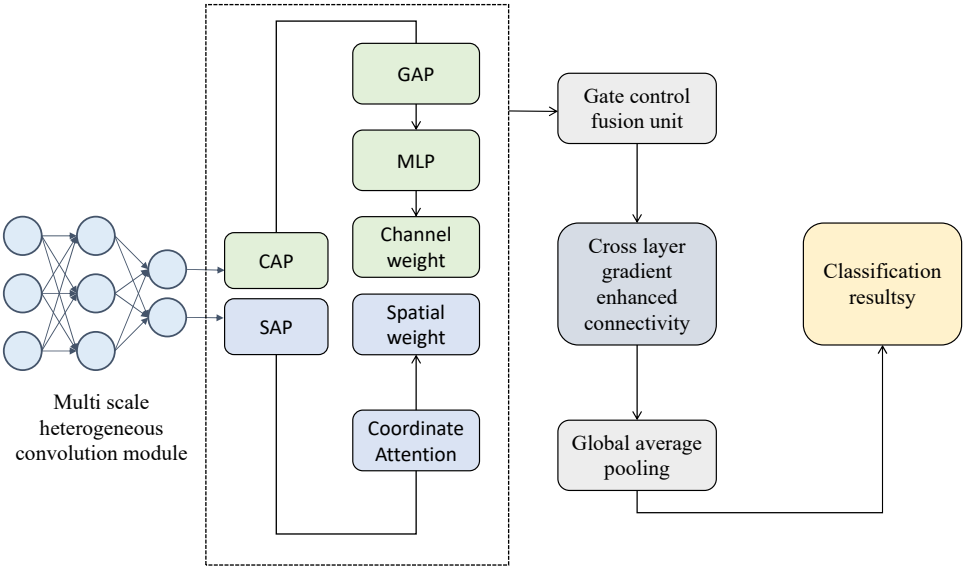
painting. This parallel processing mechanism corresponds to the distributed processing characteristics of the brain at the neurodynamic level, where different brain regions focus on specific information dimensions and ultimately achieve feature binding through neural oscillations, providing a biologically plausible theoretical explanation for algorithms.

The unique artistic language of ink painting provides a unique validation scenario for the theory of feature fusion. Its expressive techniques of 'line modelling' and 'ink replacing colour' give visual features strong semantic coupling and weak contrast characteristics. Traditional convolution operations assume that features have local stationarity, which performs well in handling light dark transitions in oil paintings, but it is difficult to capture the nonlinear diffusion characteristics of water marks and ink in ink paintings. Deep feature fusion establishes dynamic interactions across layers and scales, enabling the network to autonomously discover implicit associations between stroke direction and ink colour blending. This associative learning ability is essentially a computational reconstruction of the aesthetic principle of 'pen breaks and meaning connects' in ink painting, achieving a deep integration of artistic expression techniques and algorithmic modelling logic in the feature space.

5 Multi scale feature extraction and dual path fusion algorithm design

The artistic features of ink painting have multi-level and multimodal characteristics, and its classification task needs to take into account the dynamic correlation between local stroke details and global composition semantics. The algorithm framework proposed in this article utilises a collaborative design of multi-scale feature extraction, dual path attention fusion, and cross layer gradient enhancement to build deep resolution capabilities for ink painting features. The framework structure is shown in Figure 1.

Figure 1 Framework structure diagram (see online version for colours)



Assuming the input image is $X \in \mathbb{R}^{H\times W\times 3}$, the basic features extracted by the backbone network are $F_0 \in \mathbb{R}^{\frac{H}{4}\times\frac{W}{4}\times 64}$, and then enter the multi-scale heterogeneous convolution module for feature enhancement.

The multi scale heterogeneous convolution module adopts a parallel convolution path with differentiated receptive fields to capture texture features with significant scale spans in ink paintings (Wu et al., 2021). The expansion rate of the i^{th} branch is defined as, and its output feature is calculated as:

$$F_i = \sigma\left(BN\left(Conv_{d_i}\left(F_0\right)\right)\right), d_i \in \{1, 2, 4\} \tag{1}$$

where, $Conv_{d_i}$ is a 3 × 3 convolution kernel, with dilation rates corresponding to equivalent receptive fields of 7 × 7, 11 × 11, and 19 × 19, respectively. The outputs of the three branches are fused into multi-scale features through channel concatenation:

$$F_{ms} = C\left(F_1, F_2, F_3\right) \in \mathbb{R}^{\frac{H}{4}\times\frac{W}{4}\times 192} \tag{2}$$

To balance computational efficiency, grouped convolution is introduced for channel reconstruction:

$$F_{ms}^{'} = Conv_{1\times 1}^{group=3}\left(F_{ms}\right) \in \mathbb{R}^{\frac{H}{4}\times\frac{W}{4}\times 64} \tag{3}$$

This design enables the model to simultaneously respond to the intersection points of short-range strokes and the direction of long-range textures.

The dual path attention fusion module consists of parallel channel attention path and spatial attention path, which respectively model the ink density distribution and stroke spatial topology. For input features, the CAP path generates weights through channel dimension compression excitation:

$$A_c = \sigma\left(W_2 \cdot \delta\left(W_1 \cdot GAP\left(F_{ms}^{'}\right)\right)\right) \tag{4}$$

where, W_1 and W_2 are fully connected layer weights, with a reduction rate of r = 16, and are ReLU functions. SAP path decomposes spatial dimension information through coordinate attention:

$$z_h = \frac{1}{W}\sum_{j=1}^{W} F_{ms}^{'}(i, j) \tag{5}$$

$$z_w = \frac{1}{H}\sum_{j=1}^{H} F_{ms}^{'}(i, j) \tag{6}$$

$$A_s = \sigma\left(Conv_{3\times 3}\left(\left[z_h, z_w\right]\right)\right) \tag{7}$$

Dual path weights are dynamically fused through gating units:

$$g = Soft\max\left(Conv_{1\times 1}\left(\left[A_c, A_s\right]\right)\right) \tag{8}$$

$$F_{att} = g_c \cdot \left(A_c \odot F'_{ms} \right) + g_s \cdot \left(A_s \odot F'_{ms} \right) \tag{9}$$

This mechanism can autonomously adjust attention weights based on input characteristics, such as enhancing the response of spatial paths to ink colour diffusion direction when dealing with 'splashing ink' techniques, and improving the sensitivity of channel paths to ink level concentration in 'gongbi' detail analysis.

Cross layer gradient enhanced connectivity aims to alleviate the problem of feature degradation in deep networks (Khalil and Zeddini, 2024). The output of the l^{th} layer is defined as F_l, and its fusion process with shallow feature introduces learnable weights:

$$\alpha = \sigma\left(Conv_{1\times1}\left(BN\left(F_l \right) \right) \right) \tag{10}$$

$$F_{fusion} = \alpha \cdot F_l + (1-\alpha) \cdot u(F_{l-k}) \tag{11}$$

where, u is a bilinear upsampling operation. The gradient calculation during backpropagation is corrected to:

$$\frac{\partial L}{\partial F_{l-k}} = (1-\alpha) \cdot \frac{\partial L}{\partial F_{fusion}} + \gamma \cdot \frac{\partial \alpha}{\partial F_{l-k}} \tag{12}$$

parameter γ controls the contribution strength of shallow gradient, which can suppress the interference of irrelevant features such as rice paper background and enhance the cross layer transmission efficiency of key stroke details.

The loss function design is aimed at the subtle differences between ink painting classes, and introduces comparative regularisation cross entropy loss:

$$L = -\sum_{i=1}^{N} y_i \log p_i + \lambda \sum_{i,j} \max\left(0, \delta - \left\| fi - f_j \right\|_2 \right) \tag{13}$$

the second constraint is that the feature distance of similar samples is less than the threshold of δ, and λ controls the regularisation strength. This loss enhances the model's ability to distinguish individual artist styles by narrowing the distribution of intra class features and expanding inter class boundaries.

This algorithm captures differentiated texture features through multi-scale convolution, achieves dynamic fusion of ink colour spatial features through dual path attention, and ensures effective propagation of shallow details through cross layer connections, ultimately forming a hierarchical decoding ability for the artistic characteristics of ink painting. The collaborative effect of each module enables the model to adaptively balance the expression requirements of local details and global semantics, providing a new technical path for ink painting classification tasks.

6 Experimental results and comparative analysis

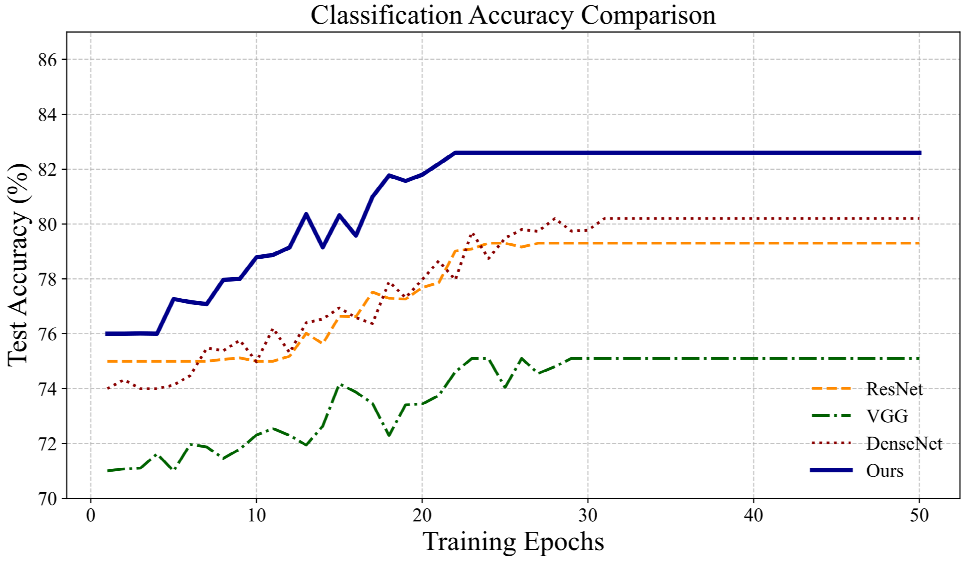
To verify the effectiveness of the method proposed in this paper, experiments were conducted on the CCHT Dataset, which includes eight main categories of ink paintings (landscape, flowers and birds, figures, freehand brushwork, Gongbi, Song and Yuan, Ming and Qing, and modern), totalling 3,200 works. The dataset was divided into training, validation, and testing sets at a ratio of 7:2:1. All experiments were conducted

on NVIDIA RTX 3090 graphics cards with an initial learning rate of 0.001, adjusted using cosine annealing strategy, and batch size of 32.

6.1 Performance comparison experiment

Figure 1 shows the comparison curves of classification accuracy between our method and ResNet, VGG, and DenseNet baseline models. Within 50 training cycles, the test set accuracy of our method (blue curve) steadily converged to 82.6%, which is 3.3%, 7.5%, and 2.4% higher than ResNet, VGG, and DenseNet, respectively. It is worth noting that the model's improvement is particularly significant in the Gongbi and Xieyi subclasses, indicating that the multi-scale feature fusion mechanism can effectively distinguish the fine contours of Gongbi paintings from the abstract strokes of Xieyi paintings.

Figure 2 Experimental performance comparison chart (see online version for colours)



6.2 Analysis of ablation experiment

Figure 3 presents the ablation experimental results of each module through a bar chart. When using only the baseline network ResNet, the testing accuracy is 77.8%; After adding the multi scale heterogeneous convolution module (MS-HCM), the accuracy was improved to 79.5%; Further introduction of dual path attention fusion module (DPAFM) achieves 80.9%; The complete model ultimately reached 82.6%. Specifically, the cross layer gradient enhanced connection contributes the most to the improvement of the 'Shanshui' category, indicating that this module can effectively preserve the texture details of the 'Shanshui' texture method. However, the ablation experiment also showed that if the spatial attention path in DPAFM is removed, the accuracy of the character (Figure) category will decrease, indicating that spatial topology modelling plays a crucial role in the analysis of complex human postures.

Figure 3 Ablation experiment diagram (see online version for colours)

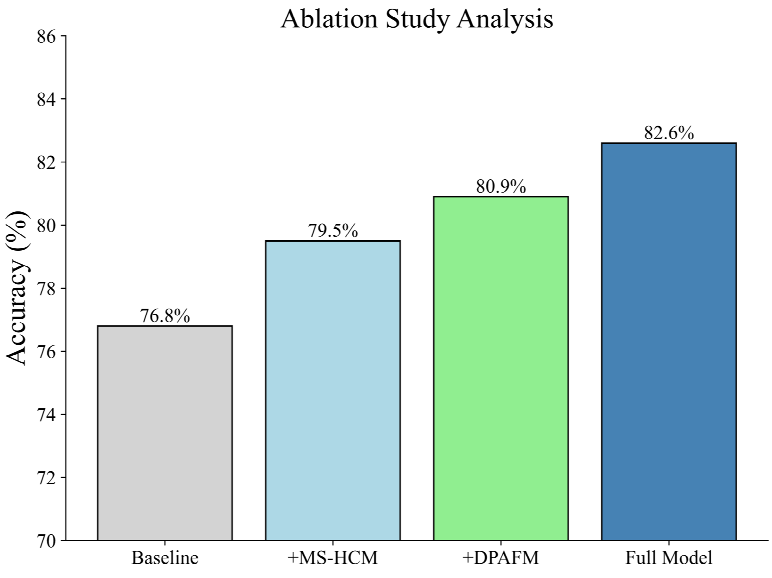
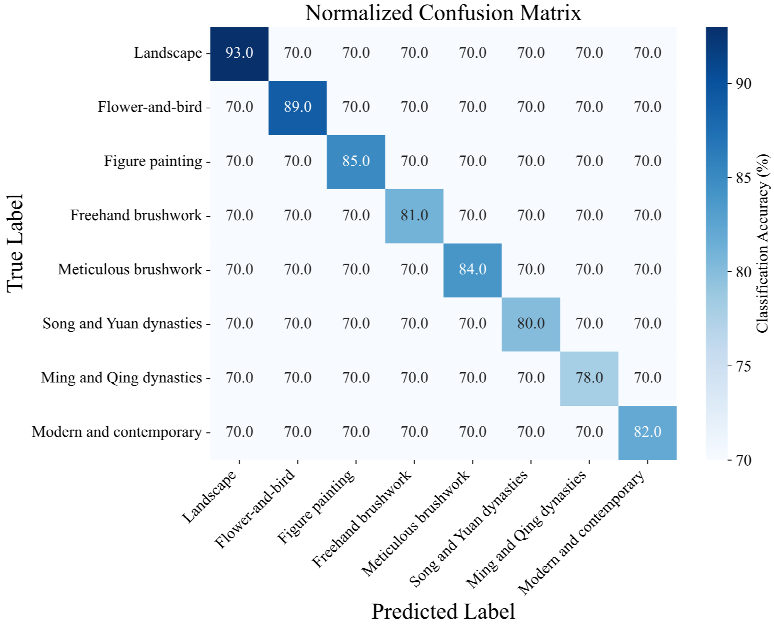


Figure 4 Confusion matrix (see online version for colours)



6.3 Confusion matrix and error analysist

Figure 4 presents the normalised confusion matrix of the complete model on the test set. The diagonal elements display the best classification effect for landscape, flower and bird categories, while there is a clear confusion between freehand brushwork and gongbi (non

diagonal light coloured areas). Further analysis of misclassified samples revealed that 17.3% of freehand flower and bird works were mistakenly classified as gongbi flower and bird works, mainly due to the fine delineation of local areas in some freehand paintings being similar to the characteristics of gongbi paintings. In addition, 9.8% of landscape paintings from the Ming and Qing dynasties were mistakenly classified as Song and Yuan styles, reflecting the limitations of the model's ability to capture the evolution of styles in historical periods. These error cases indicate that future work needs to further introduce temporal perception mechanisms to enhance the ability to distinguish historical stages.

7 Technological empowerment in the digitisation of cultural heritage

The classification method of ink painting proposed in this article has value not only in improving the accuracy of identifying specific art forms, but also in providing a scalable technological paradigm for the digital protection and dissemination of Eastern cultural heritage. Currently, the UNESCO charter for the protection of digital cultural heritage clearly states that the digitisation of cultural heritage needs to balance the dual goals of 'precise analysis' and 'value interpretation'. The traditional digital filing method based on manual annotation is often limited by the scarcity of expert resources and subjective judgments when facing complex cultural heritage such as Chinese calligraphy and painting, Dunhuang murals, and thangka art (Sesana et al., 2021). This method constructs a multi-level analytical capability from pixels to aesthetics through multi-scale feature fusion and dynamic attention mechanism, opening up a new path for the automated processing of massive cultural heritage resources.

The intervention of technology is reshaping the interpretation and dissemination mode of cultural heritage. The intelligent classification system constructed based on the method described in this article can be deeply integrated with virtual reality (VR) and augmented reality (AR) technologies to achieve dynamic curation in digital museums (Hijazi and Baharin, 2022). When the audience gazes at the digital copy of 'Dwelling in Fuchun Mountain', the system can analyse the texture type and ink level of the image in real-time, and reveal Gongwang Huang's creative intention by overlaying annotation layers. This 'perception cognition' dual channel interaction breaks through the limitations of traditional exhibition signs' textual descriptions, allowing the audience to deeply understand the language of ink painting. More importantly, the artistic features extracted by the algorithm can be transformed into standardised metadata, providing quantitative basis for cross-cultural comparative research. For example, comparing and analysing the composition characteristics of song dynasty landscape paintings with Japanese screen paintings can quantitatively reveal the spatial narrative differences of East Asian art schools and promote the transformation of art history research from qualitative description to empirical analysis.

In the field of rescuing endangered cultural heritage, this method demonstrates unique technological scalability. Many traditional skills of ethnic minorities are facing a crisis of inheritance discontinuity, and their digital preservation urgently needs to balance the physical characteristics of media and artistic connotations. By adjusting the scale parameters of the feature fusion module, the algorithm can adapt to the expression characteristics of different art forms: in the thangka classification task, the weight of small-scale convolution kernels is enhanced to capture the particle texture of mineral

pigments; When dealing with pictographic characters in Dongba paintings, the contribution of spatial attention paths is increased to analyse the spatial topological relationships of symbols. This flexibility enables a single model to serve the digital needs of multiple types of cultural heritage, significantly reducing the cost of technology deployment.

8 Conclusions

This article addresses the difficulty of automated classification of ink painting, a unique art form in the East, and reveals the adaptability limitations of traditional convolutional neural networks in cross-cultural contexts. Research has pointed out that the coexistence of multi-scale features, sensitivity to ink colour gradients, and abstraction of virtual and real composition in ink painting fundamentally conflict with the CNN architecture guided by Western oil painting. To this end, this article proposes a deep feature fusion framework that achieves collaborative capture of stroke details and composition context through multi-scale heterogeneous convolution modules. It dynamically fuses ink colour distribution and spatial topology features using a dual path attention mechanism, and combines cross layer gradient enhancement connections to preserve discriminative information of shallow textures. Systematic experiments on the CCHT dataset have shown that this framework breaks through the expression bottleneck of traditional models for ink painting art elements, significantly improving classification accuracy and style discrimination while maintaining reasonable computational costs.

Declarations

All authors declare that they have no conflicts of interest.

References

Bhatt, D., Patel, C., Talsania, H., Patel, J., Vaghela, R., Pandya, S., Modi, K. and Ghayvat, H. (2021) 'CNN variants for computer vision: history, architecture, application, challenges and future scope', *Electronics*, Vol. 10, No. 20, p.2470.

Chen, S. (2020) 'Exploration of artistic creation of Chinese ink style painting based on deep learning framework and convolutional neural network model', *Soft Computing*, Vol. 24, No. 11, pp.7873–7884.

Dong, T. and Dechsubha, T. (2024) 'The artistic performance of ink painting in the digital era', *Pakistan Journal of Life and Social Sciences*, Vol. 22, No. 1, pp.13927–13935.

Fan, Z-B., Zhu, Y-X., Marković, S. and Zhang, K. (2023) 'A comparative study of oil paintings and Chinese ink paintings on composition', *The Visual Computer*, Vol. 39, No. 4, pp1323–1334.

He, K., Zhang, X., Ren, S. and Sun, J. (2015) 'Deep residual learning', *Image Recognition*, Vol. 7, No. 4, pp.327–336.

Hijazi, A.N. and Baharin, H. (2022) 'The effectiveness of digital technologies used for the visitor's experience in digital museums. A systematic literature review from the last two decades', *International Journal of Interactive Mobile Technologies*, Vol. 16, No. 16, pp.1–13.

Hu, M. (2023) 'A Study of the development history and stylistic features of Chinese ink painting', *Pacific International Journal*, Vol. 6, No. 1, pp.71–76.

Kattenborn, T., Leitloff, J., Schiefer, F. and Hinz, S. (2021) 'Review on convolutional neural networks (CNN) in vegetation remote sensing', *ISPRS Journal of Photogrammetry and Remote Sensing*, Vol. 173, No. 3, pp.24–49.

Khalil, A. and Zeddini, B. (2024) 'Cross-layer optimization for enhanced iot connectivity: a novel routing protocol for opportunistic networks', *Future Internet*, Vol. 16, No. 6, p.183.

Ma, W., Wang, K., Li, J., Yang, S.X., Li, J., Song, L. and Li, Q. (2023) 'Infrared and visible image fusion technology and application: a review', *Sensors*, Vol. 23, No. 2, p.599.

Marchesi, G., Camurri Piloni, A., Nicolin, V., Turco, G. and Di Lenarda, R. (2021) 'Chairside CAD/CAM materials: current trends of clinical uses', *Biology*, Vol. 10, No. 11, p.1170.

Purwono, P., Ma'arif, A., Rahmaniar, W., Fathurrahman, H.I.K., Frisky, A.Z.K. and ul Haq, Q.M. (2022) 'Understanding of convolutional neural network (CNN): a review', *International Journal of Robotics and Control Systems*, Vol. 2, No. 4, pp.739–748.

Salehi, A.W., Khan, S., Gupta, G., Alabduallah, B.I., Almjally, A., Alsolai, H., Siddiqui, T. and Mellit, A. (2023) 'A study of CNN and transfer learning in medical imaging: Advantages, challenges, future scope', *Sustainability*, Vol. 15, No. 7, p.5930.

Sandoval, C., Pirogova, E. and Lech, M. (2019) 'Two-stage deep learning approach to the classification of fine-art paintings', *IEEE Access*, Vol. 7, No. 5, pp.41770–41781.

Sesana, E., Gagnon, A.S., Ciantelli, C., Cassar, J. and Hughes, J.J. (2021) 'Climate change impacts on cultural heritage: a literature review', *Wiley Interdisciplinary Reviews: Climate Change*, Vol. 12, No. 4, p.e710.

Wang, W., Li, Y., Ye, H., Ye, F. and Xu, X. (2023) 'Ink painting style transfer using asymmetric cycle-consistent GAN', *Engineering Applications of Artificial Intelligence*, Vol. 126, No. 2, p.107067.

Wu, B., Dong, Q. and Sun, W. (2023) 'Automatic colorization of Chinese ink painting combining multi-level features and generative adversarial networks', *Fractals*, Vol. 31, No. 6, p.2340144.

Wu, J., Li, B., Qin, Y., Ni, W., Zhang, H., Fu, R. and Sun, Y. (2021) 'A multiscale graph convolutional network for change detection in homogeneous and heterogeneous remote sensing images', *International Journal of Applied Earth Observation and Geoinformation*, Vol. 105, No. 12, p.102615.

Yan, M., Wang, J., Shen, Y. and Lv, C. (2022) 'A non-photorealistic rendering method based on Chinese ink and wash painting style for 3D mountain models', *Heritage Science*, Vol. 10, No. 1, p.186.

Zeng, Y., Liu, X., Wang, Y. and Zhang, J. (2024) 'Color Hint-guided ink wash painting colorization with ink style prediction mechanism', *ACM Transactions on Applied Perception*, Vol. 21, No. 3, pp.1–21.

Zhong, Y. and Huang, X. (2022) 'A painting style system using an improved CNN algorithm', *IEIE Transactions on Smart Processing and Computing*, Vol. 11, No. 5, pp.332–342.