



International Journal of Information and Communication Technology

ISSN online: 1741-8070 - ISSN print: 1466-6642

<https://www.inderscience.com/ijict>

Digital reconstruction method of intangible cultural heritage painting integrating MSI and improved U-Net

Bin Chen

DOI: [10.1504/IJICT.2026.10077189](https://doi.org/10.1504/IJICT.2026.10077189)

Article History:

Received:	13 October 2025
Last revised:	15 December 2025
Accepted:	18 December 2025
Published online:	01 April 2026

Digital reconstruction method of intangible cultural heritage painting integrating MSI and improved U-Net

Bin Chen

School of Arts,
Minnan Normal University,
Zhangzhou, 363000, China
Email: xwz7401@163.com

Abstract: Intangible cultural heritage paintings often suffer from fading, structural damage, and detail loss due to environmental and material aging, making high-fidelity digital reconstruction essential for cultural preservation. To address the limitations of red-green-blue imaging (RGB) and insufficient spectral-spatial modelling, this study proposes context-aware multi-spectral imaging network (CA-MSI-Net), a reconstruction method integrating multi-spectral imaging (MSI) with an improved U-Net architecture. Spatial and channel transformer modules are embedded to enhance long-range spatial dependencies and cross-band spectral interactions, while contextual modelling and multi-scale attention mechanisms strengthen texture perception and boundary restoration. Experiments on multispectral datasets demonstrate that CA-MSI-Net achieves superior reconstruction accuracy, with mean intersection over union (mIoU), dice similarity coefficient (DSC), and F1 score reaching 81.3%, 92.1%, and 0.94, respectively, outperforming UCTransNet and dynamic optimisation vision network (DovNet). The method shows robust performance across painting styles, lighting conditions, and spectral configurations.

Keywords: intangible cultural heritage painting; digital reconstruction; multi-scale attention mechanism; multi-spectral imaging; U-Net.

Reference to this paper should be made as follows: Chen, B. (2026) 'Digital reconstruction method of intangible cultural heritage painting integrating MSI and improved U-Net', *Int. J. Information and Communication Technology*, Vol. 27, No. 30, pp.49–72.

Biographical notes: Bin Chen obtained his Master's degree from Fujian Normal University in 2010. He currently serves as an Associate Professor and Assistant Dean at the School of Art, Minnan Normal University. He was a visiting scholar at Tsinghua University, and holds the positions of Deputy Secretary-General of the Oil Painting Art Committee of Fujian Artists Association and Executive Director of the Fujian Art Education Committee. His research focuses on Folk Art.

1 Background

As an important carrier of traditional art, intangible cultural heritage painting carries unique historical information and aesthetic value. However, due to long-term environmental erosion, material aging, and other factors, the original work often suffers from fading, peeling, and missing details. There is an urgent need for high-precision

digital reconstruction to support protection and inheritance (Zhang, 2023; Guo, 2023). Multi-spectral imaging (MSI) has the advantage of obtaining rich spectral information and has been widely used in cultural relic detection and art restoration (Liu et al., 2023; Zhang et al., 2024a). However, existing methods still have shortcomings in spectral and spatial information fusion, multi-scale detail modelling, robustness to lighting and noise, and control of high channel computational overhead, making it difficult to balance reconstruction accuracy, computational efficiency, and multi-scene adaptability. This places demanding requirements on the high-fidelity restoration of complex textures and multi-style features in intangible cultural heritage paintings (Rathi et al., 2024; Ma and Ng, 2025). At present, extensive research has been conducted on the digital reconstruction of intangible cultural heritage art both domestically and internationally.

Rihani (2023) proposed a method for reconstructing ancient temples based on crowdsourcing images and 3D photogrammetry technology, which emphasised the construction of interactive immersive experiences and enhanced data diversity through public participation. However, this method is weak in controlling image quality and spectral consistency, making it difficult to ensure high fidelity of reconstructed details. Belhi et al. (2023) designed a machine learning framework for enhancing digital experiences of cultural heritage, which achieved digital display of cultural scenes through the integration of multi-source data and interactive applications. However, its focus was on optimising user experience, and there was insufficient guarantee for the spectral and geometric accuracy of image reconstruction. Xiang et al. (2024) conducted research on digital reconstruction of historical and cultural landscapes based on image recognition technology, achieving automatic recognition and restoration of scene elements. This method performed outstandingly in structural recognition, but did not introduce multi-spectral or hyperspectral information, resulting in limited accuracy in colour and texture restoration. Zhang and Taib (2025) proposed a virtual experimental modelling and digital protection path for traditional Huizhou wood carving, and constructed an interactive 3D virtual model. Although it had advantages in structural preservation and display, its integration and utilisation of MSI and reconstruction algorithms were insufficient, making it difficult to meet the high-precision restoration requirements of intangible cultural heritage painting details and textures.

Zhang and Cao (2025) established an off-axis digital holographic imaging framework using the fractional Fourier transform, significantly enhancing phase reconstruction accuracy and anti-aberration performance. In porous media research, Al Saad (2024) realised 3D recording and visualisation of ancient structures, offering a basis for texture restoration. Gao et al. (2025) proposed a greyscale-assisted RGB conversion method for near-infrared images, improving structural and texture visibility. These advances – from optical imaging to generative and attention-based modelling – have improved structural fidelity and reconstruction robustness, yet most methods still model spectral information only implicitly and lack cross-style adaptability. Addressing these limitations, this study targets multi-spectral reconstruction of intangible cultural heritage paintings and introduces a multi-scale, context-aware spectral-spatial joint modelling framework to ensure spectral consistency, structural integrity, and stylistic adaptability.

In contrast, the proposed method combines MSI with an improved U-Net structure to simultaneously account for spectral restoration accuracy and spatial structural details in the digital reconstruction of intangible cultural heritage paintings. The proposed CA-MSI-Net integrates MSI with an enhanced U-Net architecture to jointly improve spectral restoration and spatial structural reconstruction, overcoming limitations in colour

consistency, texture fidelity, and spectral utilisation. An MSI system with spectral decoupling provides high-resolution multi-band data, while spatial and channel transformers embedded in U-Net3+ enhance long-range cross-band modelling and semantic understanding. Furthermore, multi-scale feature grouping and orientation-aware contextual attention strengthen the reconstruction of complex textures and boundary details. The innovation of this research lies in establishing a spectral-spatial collaborative modelling paradigm. CA-MSI-Net links multi-band physical imaging with semantic reconstruction, providing a unified framework for interpreting pigment layers, structural patterns, and stylistic variations in cultural artworks. Through multi-scale context awareness and directional feature modelling, it offers new insights into artistic expression in digital restoration, while its lightweight design supports deployment in museum archival systems and mobile cultural platforms. The research provides foundational digital data resources for museums, archives, and cultural tourism institutions. It also supports the development of the cultural and creative industries and helps reduce preservation risks associated with the physical exhibition of original artworks.

2 Methodology

2.1 MSI feature extraction method based on improved U-Net3+

Within the protection system of intangible cultural heritage, painting works are not only artistic expressions of aesthetics but also visual symbols of historical memory and social identity. Styles such as gongbi, murals, and silk paintings are rooted in specific historical periods and regional cultures, embodying both traditional craftsmanship and the externalisation of collective cultural psychology. Yet their preservation faces increasing pressure: environmental changes and light exposure cause fading and structural damage, while the decline of inheritors and the weakening of oral transmission place these traditions at risk of distortion and loss. Digital reconstruction thus becomes not only a technical remedy but also a cultural preservation strategy. Through high-precision archives and visual restoration platforms, it helps break temporal and spatial barriers, supporting cross-generational and cross-regional transmission.

In the digital reconstruction of intangible cultural heritage paintings, obtaining high-quality MSI data is a prerequisite for achieving detail fidelity, colour restoration, and texture analysis (Mu et al., 2021; Lei et al., 2023). The traditional RGB imaging method suffers from severe information loss in the colour dimension, making it difficult to meet the structural restoration requirements in complex painting scenes (Chen et al., 2022; Zhang et al., 2024b). An MSI acquisition system with spectral decoupling capability is adopted, consisting of a liquid crystal tunable filter (LCTF, 400–1000 nm, FWHM $\approx$ 10 nm) and a scientific CMOS sensor (FLIR GS3-U3-51S5M-C, $2,448 \times 2,048$ pixels). Multi-band images under a single exposure are obtained by combining sparse spectral coding, spatial modulation, and filter sequence switching. A calibrated 3,200 K halogen light source ensures stable illumination, with the average SNR maintained above 45 dB to reduce spectral interference. Hardware synchronisation via a USB3.0 high-speed interface guarantees real-time spectral alignment. Spectral decoupling capability means that the imaging system can independently extract and construct the reflection signals of different bands into multi-channel data through spectral filtering or spatial modulation. This enables the realisation of the physical layer

distinction of each spectral component in the acquisition stage. It also avoids colour mixing errors caused by spectral overlap in traditional RGB imaging. The schematic diagram of MSI acquisition system is shown in Figure 1.

Figure 1 Schematic diagram of MSI acquisition system (see online version for colours)

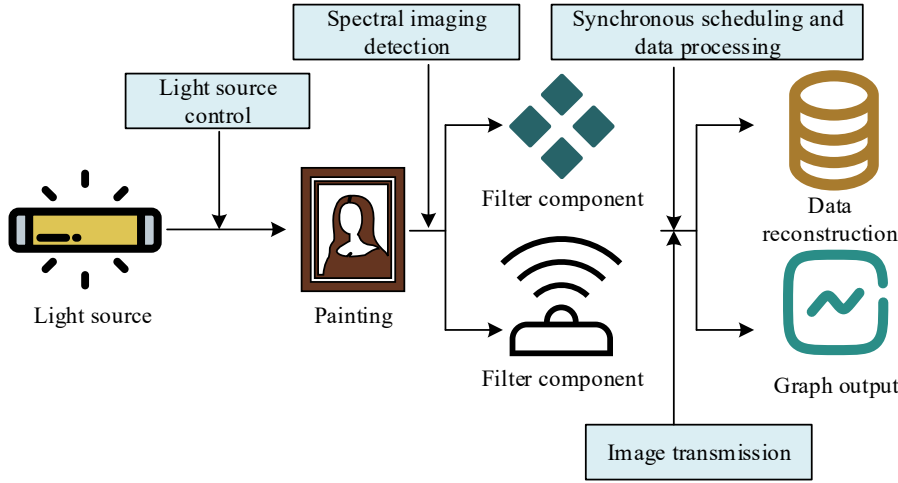
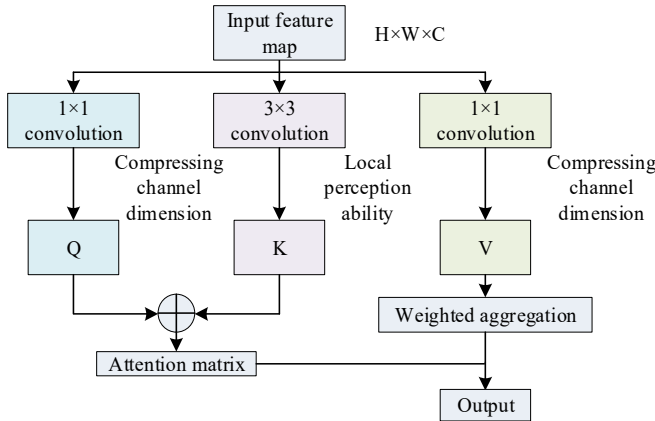


Figure 2 Schematic diagram of spatial attention module structure in improved U-Net3+network (see online version for colours)



In Figure 1, the MSI acquisition system integrates multiple modules such as light source control, spectral imaging detection, image transmission, synchronous scheduling, and data processing. This system can achieve non-contact acquisition of target images in multiple bands and high resolution without damaging the original painting. In the actual acquisition process, the system uses a highly stable photon source to fully illuminate the sample, and obtains reflection signals in multiple spectral bands through filter components and spectral image sensors. Finally, the image control and visual processing modules complete data reconstruction and spectral output (Shafique et al., 2023; Lee et al., 2023). This system has the advantages of wide spectral coverage, high collection efficiency, and low information loss, and is suitable for obtaining raw data of complex

pigments, fine textures, and hierarchical structures in intangible cultural heritage painting. After the initial acquisition and structural reconstruction of multi-band images, an MSI deep reconstruction network based on an improved U-Net3+ is further constructed. To enhance the ability to capture long-range dependent features in intangible cultural heritage images, the U-Net3+architecture is improved by introducing a spatial transformer module in the backbone U-Net3+structure. This module replaces the simple concatenation method between low order and high-order features in traditional skip connections (Duarte et al., 2025; Park et al., 2025). In this module, a self-attention mechanism is introduced to encode and model the interactions between input features, including queries (Q), keys (K), and values (V), in order to achieve dynamic fusion and reconstruction of contextual information between pixels within a spatial range. Its specific structure is shown in Figure 2.

In Figure 2, the spatial transformer module is embedded in the skip connection channel of U-Net3+to enhance the contextual modelling ability in the spatial dimension of MSI images. The spatial transformer module is configured with an input feature map size of $H \times W \times C_1$ ($256 \times 256 \times 64$). After positional encoding and projection, the feature map is flattened to $(H \times W) \times d_1$ ($65,536 \times 64$). A 4-head self-attention mechanism (head dimension 16) is utilised to jointly model long-range dependencies. The output tensor is reshaped back to $H \times W \times C_1$, and fused with the original feature by residual connection, resulting in no change in spatial resolution. The input feature map extracts query features (Q), key features (K), and value features (V) through three different paths. The key features are generated by a 3×3 convolution kernel to preserve local perceptual ability, while the query features and value features are generated by a $1 \times$ convolution to compress the channel dimension and improve computational efficiency. In the key value matching process, query features are concatenated with key features and mapped into an attention matrix, which is used to characterise the spatial relationship between the current pixel and its surrounding local area. By weighting and aggregating the value features within the neighbourhood, the model dynamically adjusts the spatial expression of each position to obtain the reconstructed features. Finally, the enhanced feature is fused with the original input through residual connections, preserving local texture information and improving overall spatial consistency. The MSI image input representation and feature extraction are shown in equation (1).

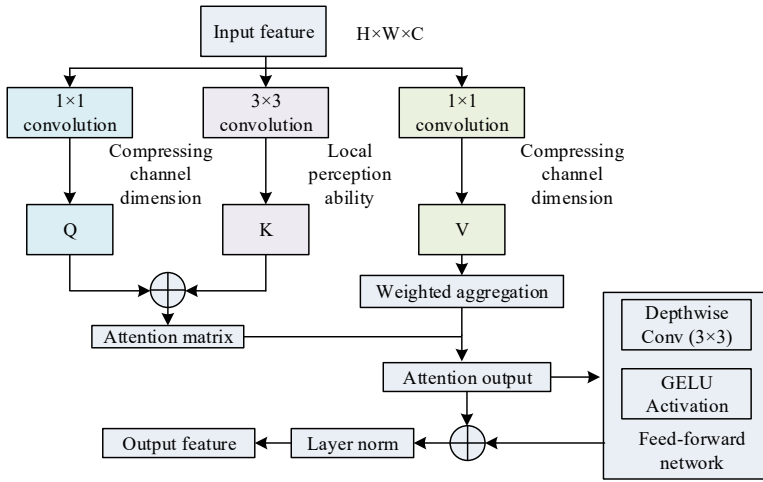
$$\begin{cases} X \in \mathbb{R}^{H \times W \times C} \\ F^{(0)} = \phi(X) \end{cases} \quad (1)$$

In equation (1), X is the input MSI image, an H , W , and C represent the height, width, and number of spectral channels of the image, respectively. $\phi(\cdot)$ is the feature mapping function that represents the combination of linear transformations and nonlinear functions. $F^{(0)}$ is the deep feature representation of initial extraction. The calculation process of spatial self attention mechanism is shown in equation (2).

$$\begin{cases} Q = W_Q F \\ K = W_K F \\ V = W_V F \\ A = \frac{Q K^T}{\sqrt{d}} \\ S = \sigma(A) \\ F^{(1)} = S V \end{cases} \quad (2)$$

In equation (2), F represents the two-dimensional spatial feature, W is the weight matrix, A is the attention score matrix, $\sigma(\cdot)$ is the softmax function, and $F^{(1)}$ is the output feature after spatial attention. T is the transpose of matrix K , d is the dimension of each query vector, and S is the normalised attention weight matrix. To further enhance the feature expression ability of the model in the spectral dimension, a channel dimension feature modelling module is introduced in the improved U-Net3+, as shown in Figure 3.

Figure 3 Schematic diagram of channel dimension feature modelling module structure (see online version for colours)



In Figure 3, the channel dimension feature modelling module explicitly models the channel dimension of MSI images through multi-head attention mechanism to capture semantic dependencies and mutual influences between different bands. The input features are first mapped into query, key, and value vectors, and after attention weighting in the channel dimension, fused features are output. Subsequently, the information between channels is further extracted through stacked channel transformer blocks and feed-forward neural networks to enhance the consistency of colour structure expression. The feed-forward neural network module introduces depthwise separable convolution and GELU activation function, which enhances nonlinear modelling capability while maintaining lightweight. For the channel transformer, the input tensor is resized to $C \times (H \cdot W)$ ($12 \times 65,536$), where C is the number of spectral channels. A 6-head attention mechanism (head dimension 32) is adopted, allowing group-based feature interactions across spectral channels. Following feed-forward fusion and depthwise

separable convolution, the output size remains $H \times W \times C_2$ (consistent with input), ensuring compatibility with downstream blocks. This module enhances the network's ability to restore spectral variations such as colour transitions and pigment layering in intangible cultural heritage paintings. Compared with traditional attention mechanisms, spatial transformer explicitly models long-distance dependencies across pixel areas through self-attention structures, and can better adapt to the large-scale texture changes and complex structural features present in intangible cultural heritage paintings. The channel-dimensional feature modelling module uses a multi-head self-attention mechanism to realise group modelling and feature interaction between different spectral channels, allowing the model to perform fine-grained expressions of pigment level changes and spectral domain coupling features. This mechanism processes different spectral domain feature combinations in parallel through multiple attention heads, and performs cascade integration in the feature fusion stage, which helps improve spectral consistency recovery capabilities. The spectral channel dimension attention modelling is shown in equation (3).

$$\begin{cases} Q_c = F_c W_Q^{(c)} \\ K_c = F_c W_K^{(c)} \\ V_c = F_c W_V^{(c)} \\ A_c = \frac{Q_c K_c^T}{\sqrt{d}} \\ S_c = \sigma(A_c) \\ F^{(2)} = S_c V_c \end{cases} \quad (3)$$

In equation (3), F_c represents the feature matrix after spectral dimension rearrangement, $W^{(c)}$ represents the projection matrix of spectral domain attention, A_c represents the channel dimension attention score, S_c represents the spectral weight matrix, and $F^{(2)}$ represents the output features of channel attention. The feature fusion and residual connection are shown in equation (4).

$$F^{(3)} = F^{(0)} + F^{(1)} + F^{(2)} \quad (4)$$

In equation (4), $F^{(3)}$ is the context enhanced feature after fusion. The image reconstruction output is shown in equation (5).

$$\hat{X} = \psi(F^{(3)}) \quad (5)$$

In equation (5), $\hat{X}$ represents the reconstructed image output by the model. Based on the above modelling mechanism, a complete improved U-Net3+network structure was constructed, and its overall architecture is shown in Figure 4.

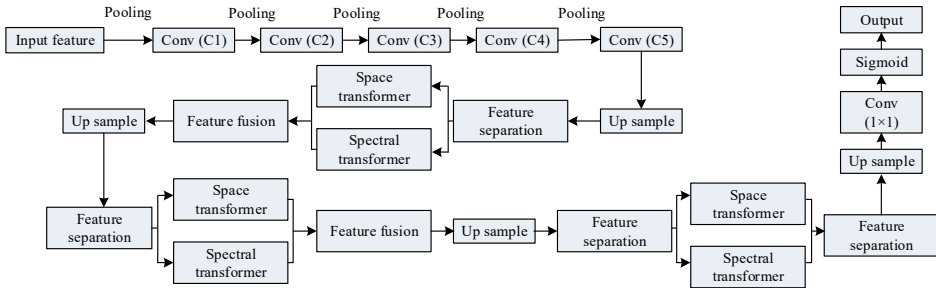
In Figure 4, the improved U-Net3+network model is extended based on U-Net3+, using an encoding decoding architecture and a dense skip connection strategy to achieve multi-scale information interaction and fusion. In the encoding stage, the original MSI image is downsampled layer by layer through multiple convolutional blocks to generate feature maps of different resolutions, and local structures and primary semantic information are extracted. Additionally, the network establishes cross layer connections between each level of scale to utilise shallow textures and deep semantics for

collaborative modelling. In the skip connection path, parallel spatial transformer and spectral transformer sub modules are introduced. The former is used to capture contextual relationships between image spatial regions, while the latter is used to establish long-range dependencies between different bands to achieve spatial and spectral synergistic enhancement. In the decoding stage, the network gradually upsamples and fuses multi-scale features from the encoding end to restore the original resolution of the image. The decoding module refines high-level abstract features through convolution operations and combines enhanced information from skip paths to improve the edge clarity and spectral consistency of the output image. Normalisation and lightweight output modules are introduced at the tail of the network to adjust channel distribution and optimise model parameter size, enhancing the system's adaptability in edge devices or deployment scenarios. The definition of the total loss function L is shown in equation (6).

$$L = \lambda_1 \cdot \|\hat{X} - X\|_1 + \lambda_2 \cdot \left(1 - \frac{2\mu_{\hat{X}}\mu_X + C_1}{\mu_{\hat{X}}^2 + \mu_X^2 + C_1} \cdot \frac{2\sigma_{\hat{X}X} + C_2}{\sigma_{\hat{X}}^2 + \sigma_X^2 + C_2} \right) \quad (6)$$

In equation (6), $\mu_{\hat{X}}, \mu_X$ are the means of the predicted image and the original image, $\sigma_{\hat{X}}^2, \sigma_X^2$ are the variances, and $\sigma_{\hat{X}X}$ is the covariance. C_1 and C_2 are constants. λ_1 and λ_2 are loss weighting factors.

Figure 4 Improved U-Net3+ network architecture diagram (see online version for colours)

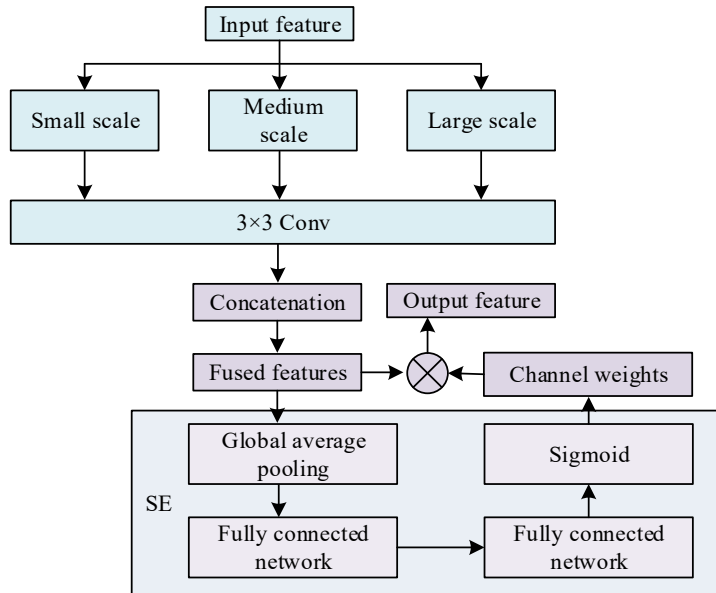


2.2 CoT-U-Net image reconstruction network based on context modelling and multi-scale fusion

In the above MSI image reconstruction method based on improved U-Net3+, the research focuses on collaborative enhancement and deep semantic modelling in spatial and channel dimensions, in order to achieve structural perception and detail restoration of intangible cultural heritage painting images. However, when facing image blocks with significant interference between complex texture regions and channel information, the network still suffers from insufficient local expression ability and redundant feature channels (Li et al., 2023; Wu et al., 2023). Therefore, a fusion reconstruction strategy based on CS technology and context modelling is further introduced, which enhances feature representation while maintaining a lightweight network, thereby improving the resolution and hierarchical structure of images in the reconstruction process (Valdez et al., 2024; Zhao et al., 2024). Among them, ‘context modelling’ is used to capture potential semantic associations between different areas of the image. By enhancing the

consistency of the expression of features in spatial structure and spectral change trends, the network can achieve more accurate information fusion in complex texture areas. On the basis of the encoding and decoding structure, the CoT-U-Net structure incorporating CS ideas and context transformer mechanism is studied, and a lightweight and high-performance reconstruction module is constructed by combining multi-scale feature grouping and channel attention mechanism (Nouasria and Et-tolba, 2022; Banu and Deivalakshmi, 2023; Pal et al., 2023). The key module structure is shown in Figure 5.

Figure 5 Structure diagram of multi-scale attention enhancement module (see online version for colours)



In Figure 5, the multi-scale attention enhancement module divides the input features into multiple subgroups and sequentially achieves cross scale feature cascading through a shared 3×3 convolution kernel. At the end of the module, a squeeze and excitation (SE) mechanism is introduced to achieve weighted modelling of channel importance in the fused feature map, further suppressing redundant features. The SE module structure consists of global average pooling, two-layer fully connected networks, and activation functions. This module generates a set of channel weight coefficients and multiplies them with the original features channel by channel, dynamically adjusting the response strength of each channel to improve the selectivity and discriminability of feature expression. To further improve the performance of the model in spatial structure modelling, a direction aware context self attention mechanism was introduced in the multi-scale attention enhancement module. This mechanism aims to capture the dependency relationships between different spatial directions in MSI images and enhance the expressive power of image structural information. Its structure is shown in Figure 6.

In Figure 6, the bidirectional attention modelling mechanism that integrates direction dependence independently models image features along the x-axis and y-axis spatial directions, and dynamically enhances direction perception through contextual interaction operations. In this scenario, context interaction operation refers to establishing spatial

structure dependence along the x-axis and y-axis directions and strengthening the feature reasoning process through directional attention. It is suitable for the recovery of linear texture and stroke direction in intangible cultural heritage paintings. In each direction, the input features are first mapped into three sets of vectors: query, key, and value. The query and key are generated through a 3×3 convolution to capture local context, while the value vector is generated through a 1×1 convolution to compress the channel dimension. On this basis, an attention weight matrix is generated through inner product operation to measure the correlation between the target position and its directional neighbourhood. Subsequently, the value vectors are weighted and summed to obtain enhanced feature representations. The attention features from both directions are fused with the original input features using residual fusion, and finally concatenated to form the final output. To clarify the interaction between directional dependency modelling and multi-scale fusion, direction-aware context features are computed independently within each spatial-resolution branch of the multi-scale grouping stage. Features from low-, mid-, and high-resolution pathways are first refined by the multi-resolution attention unit, then sequentially fed into the direction-aware module. This module models spatial dependencies along the x- and y-axes at each scale, and the resulting directional features are aggregated through weighted summation. The grouped features are then re-concatenated and passed through a shared fusion convolution layer, enabling coordinated enhancement of structural continuity and directional texture recovery. This process ensures consistent integration of global structure and directional texture evolution across multiple resolution levels. To achieve digital reconstruction of intangible cultural heritage paintings, CA-MSI-Net is developed, and its overall architecture is shown in Figure 7. This network integrates multi-scale convolution, contextual attention modelling, and directional enhancement mechanisms, taking into account the hierarchical expression of intangible cultural heritage painting images and the consistent restoration of spectral patterns.

Figure 6 Bi-directional attention modelling structure with fusion direction dependence (see online version for colours)

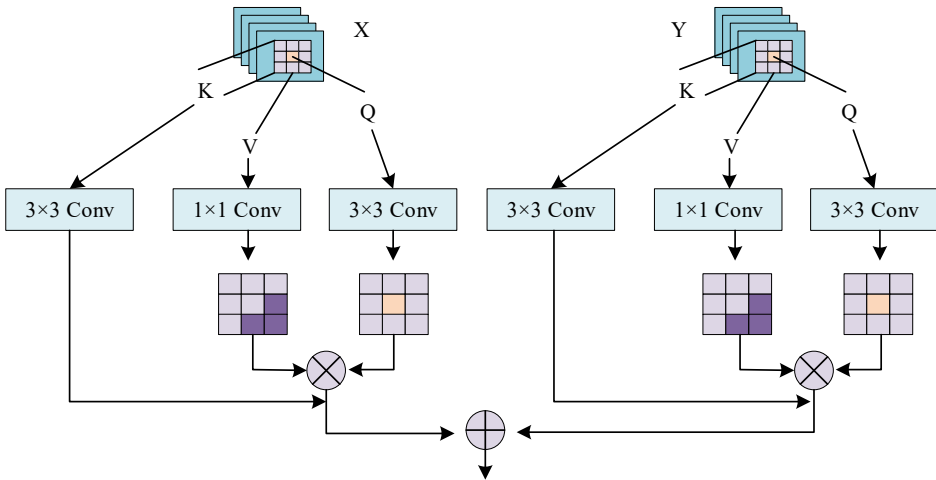
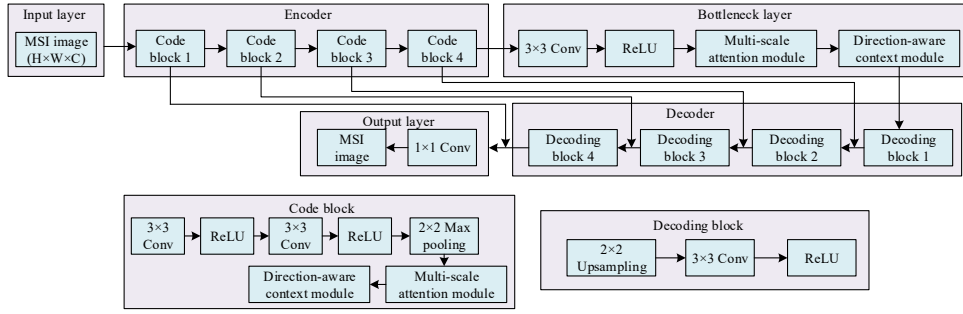


Figure 7 Schematic diagram of CA-MSI-Net structure (see online version for colours)

In Figure 7, the CA-MSI-Net adopts an encoding decoding architecture. In the encoding stage, image features are sequentially processed through two 3×3 convolutional blocks to extract basic semantic and local texture information, and spatial compression is completed through 2×2 max pooling. At the end of each encoding layer, a multi-scale attention module and a direction aware context modelling module are embedded to enhance feature expression ability and structure aware depth. The decoding part uses 2×2 upsampling for spatial recovery and skip connections with the corresponding encoded feature maps to achieve multi-scale semantic fusion and detail restoration. After each level of upsampling, convolutional blocks are introduced to refine the reconstructed expression, and finally the reconstructed image is output through 1×1 convolution. During the encoding stage of the CA-MSI-Net, the input image undergoes multiple layers of convolution and downsampling to extract multi-scale features. The output of the encoding layer is denoted as equation (7).

$$F^{(l)} = D^{(l)} \left(\sigma \left(W_c^{(l)} * F^{(l-1)} + b^{(l)} \right) \right) \quad (7)$$

In equation (7), $W_c^{(l)}$ represents the parameters of the convolutional kernel at the l^{th} layer, $b^{(l)}$ represents the bias term, $*$ denotes the convolution operation, and $D^{(l)}$ represents the downsampling operation. The embedded direction aware contextual attention and multi-scale feature fusion module is represented as shown in equation (8).

$$\tilde{F}^{(l)} = A^{(l)} \left(MSA(F^{(l)}) \right) \quad (8)$$

In equation (8), $\tilde{F}^{(l)}$ represents the feature representation after context enhancement processing. $A^{(l)}$ represents a bidirectional context modelling mechanism. MSA represents a multi-scale attention processing function. This function combines local spatial neighbourhood information and context dependence in different directions for modelling, making the model more robust in terms of structural continuity and detail level expression. In the decoding path, the decoded features are concatenated and fused with the corresponding encoding layer features, resulting in the fused features shown in equation (9).

$$R^{(l)} = u^{(l)} \left(\tilde{F}^{(l+1)} \right) \oplus \tilde{F}^{(l)} \quad (9)$$

In equation (9), $u^{(l)}$ represents the upsampling operation, $\oplus$ represents the feature concatenation operation, and $R^{(l)}$ is the fused feature after decoding at the l^{th} layer. The network end obtains the final reconstructed image output through convolutional mapping, as shown in equation (10).

$$\hat{X} = W_o * R^{(0)} + b_o \quad (10)$$

In equation (10), W_o and b_o are the weights and biases of the final 1×1 convolutional layer, and $\hat{X}$ is the reconstructed MSI image. The spatial transformer is grounded in contextual feature association theory. By using self-attention to establish dependencies between non-adjacent pixels, it overcomes the receptive-field limitations of convolution and aids in restoring cross-regional textures and complex structural patterns in intangible cultural heritage paintings. The channel transformer, rooted in spectral correlation modelling theory, leverages the inherent continuity of adjacent spectral bands. Through multi-head attention in the spectral dimension, it builds a cross-band interaction matrix that enables fine-grained reconstruction of spectral-dependent information such as colour gradients and pigment stacking.

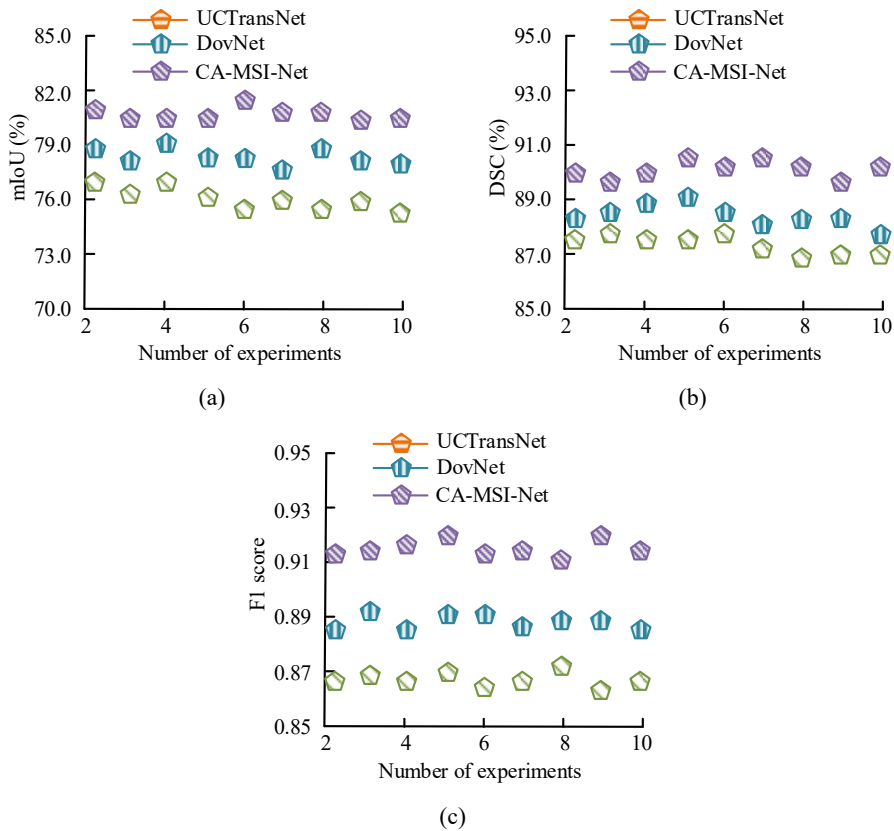
3 Results

3.1 Performance analysis of CA-MSI-netreconstruction algorithm

To verify the performance of the proposed CA-MSI-Net in the digital reconstruction of intangible cultural heritage painting images, a comparative experiment was designed, and training and testing were conducted in a unified environment to ensure the objectivity and reproducibility of the evaluation results. The experimental environment was configured as follows: the hardware platform utilised four NVIDIA A100 GPUs (each with 40 GB of memory) for parallel training, an Intel Xeon Platinum 8358 CPU (32 cores, 64 threads), and 256 GB of system memory. The software environment included the Ubuntu 20.04 operating system, the PyTorch 1.12.1 deep learning framework, and acceleration libraries such as CUDA 11.6 and cuDNN 8.4.0. The image preprocessing and data loading modules were developed using OpenCV 4.5.5 and Python 3.8. During training, mixed precision and gradient clipping mechanisms were enabled to accelerate convergence and suppress gradient explosion. Regarding dataset construction and processing, the experiment employed the CAVE Dataset and Harvard Dataset, which together comprised 3,600 sample groups covering three typical intangible cultural heritage painting styles: meticulous brushwork, murals, and silk-mounted paintings. Among the 3,600 groups of samples in the CAVE Dataset and Harvard Dataset, the meticulous samples accounted for 38.1% (1,372 groups), the mural samples accounted for 34.7% (1,250 groups), and the silk samples accounted for 27.2% (978 groups). The training set and the test set were independently divided within the style type at a ratio of 8:2 to ensure the consistency of style distribution. Each sample group included an original RGB image, a 12-channel MSI image, and a reconstruction target image annotated by experts through manual restoration. The input image size was uniformly set to $256 \times 256 \times 12$, with standardised preprocessing applied. The comparative models included an improved U-shaped network 3 plus (U-Net3+), typical transformer structure model (Swin-Transformer U-Net), a U-Net structure integrating channel-enhanced attention and transformer (UCTransNet), a

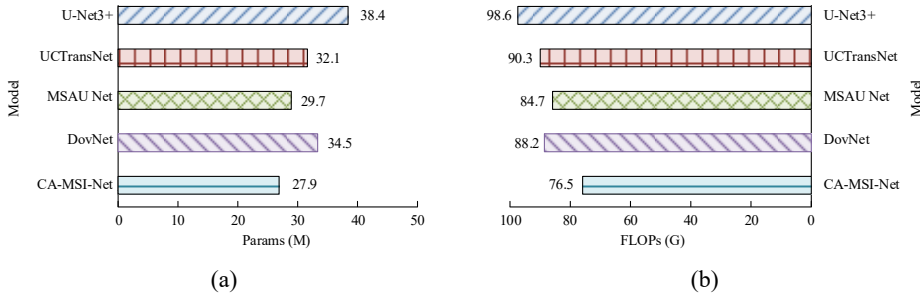
multi-scale attention U-Net (MSAU-Net), and a dynamic optimisation vision network (DovNet). U-Net3+ represents a typical encoder-decoder convolutional structure, with a wide range of applications in fields such as medical imaging and cultural relic restoration. It is structurally stable and has outstanding multi-scale feature processing capabilities. Swin-transformer U-Net and UTransNet integrate the global modelling mechanism of transformer, reflecting the trend of high-precision reconstruction tasks towards long-range dependency representation and self-attention structures in recent years. MSAU-Net introduces a multi-scale attention strategy, which has stronger local feature recognition capabilities for complex texture regions. DovNet emphasises parameter optimisation and network lightweighting, representing the algorithm development direction for engineering deployment and real-time reconstruction scenarios. The evaluation metrics adopted in the comparative experiment were mean intersection over union (mIoU), Dice similarity coefficient (DSC), F1-score, giga floating point operations (FLOPs) and learned perceptual image patch similarity (LPIPS). The performance comparison results of different methods for structural reconstruction are shown in Figure 8.

Figure 8 Comparison of structural reconstruction performance of different methods, (a) mIoU of different algorithms (b) DSC of different algorithms (c) F1-score of different algorithms (see online version for colours)



As shown in Figure 8(a), the CA-MSI-Net model consistently had a higher mean mIoU than the control model in ten experiments, with an average of 81.3%. Compared to UCTransNet (77.6%), it improved by 3.7%, and compared to DovNet (78.5%), it improved by 2.8%, indicating that it is more robust in overall structural restoration. According to Figure 8(b), the average score of CA-MSI-Net in DSC index was 92.1%, significantly better than UCTransNet (89.3%) and DovNet (90.2%), indicating that it had stronger reconstruction consistency in complex region boundary processing. As shown in Figure 8(c), the F1-score of CA-MSI-Net remained above 0.92, with the highest being 0.94, which was about 3.2% and 2.1% higher than UCTransNet and DovNet, respectively, indicating that it performed better in small target and detail structure recovery. U-Net3+ maintained stable encoding-decoding but produced blurred details in complex textures. UCTransNet improved texture modelling through channel-enhanced transformers yet lacked cross-scale structural dependency modelling. DovNet optimised convolution efficiency but underperformed in small-target recovery. In contrast, CA-MSI-Net integrated spatial context modelling, multi-scale attention, and directional enhancement, achieving synergistic gains in boundary clarity and texture reconstruction. Experimental data showed that CA-MSI-Net consistently achieved high structural reconstruction scores in multiple experiments, validating its modelling advantages of integrating attention mechanisms and multi-scale features. It is suitable for reconstruction tasks in intangible cultural heritage painting scenes that require high texture fidelity and detail restoration. The comparison results of model lightweighting indicators are shown in Figure 9.

Figure 9 Comparison results of model lightweighting indicators, (a) Params of different models (b) FLOPs of different models (see online version for colours)



As shown in Figure 9(a), in terms of parameter count, U-Net3+ had 38.4 M parameters, while DovNet and UCTransNet had 34.5 M and 32.1 M parameters, respectively. MSAU Net had the smallest parameter count at 29.7 M, whereas the proposed CA-MSI-Net had only 27.9 M parameters, the lowest among the comparative models. This indicated that the CA-MSI-Net model achieved efficient parameter utilisation in its architectural design, significantly reducing the storage and deployment burdens of the model. As depicted in Figure 9(b), regarding FLOPs, U-Net3+ had the highest value at 98.6 G, while UCTransNet and DovNet had 90.3 G and 88.2 G, respectively. MSAU-Net recorded 84.7 G, whereas CA-MSI-Net had only 76.5 G. This demonstrated that CA-MSI-Net significantly reduced the computational resources required for inference while maintaining accuracy, exhibiting stronger real-time performance and device adaptability. The experimental data revealed that the proposed CA-MSI-Net achieved optimal

lightweighting while ensuring superior performance, demonstrating good computational efficiency and model deployment potential. It was well-suited for engineering applications involving high-resolution, complex-texture tasks such as intangible cultural heritage image processing. The results of the CA-MSI-Net module ablation experiment are shown in Table 1.

Table 1 Ablation experiment

<i>Model variants</i>	<i>mIoU (%)</i>	<i>DSC (%)</i>	<i>F1-score</i>
Contextless modelling	77.4	87.0	0.86
No multi-scale attention	78.6	88.2	0.88
Undirected enhancement mechanism	79.5	89.1	0.89
Contextless + Undirected	75.6	85.2	0.83
Contextless + No multi-scale attention	76.3	86.1	0.85
No multi-scale attention + Undirected	77.1	87.5	0.86
CA-MSI-Net	81.5	90.8	0.91

Figure 10 Model stability test at different resolutions, (a) mIoU at different resolutions (b) DSC at different resolutions (c) F1-score at different resolutions (see online version for colours)

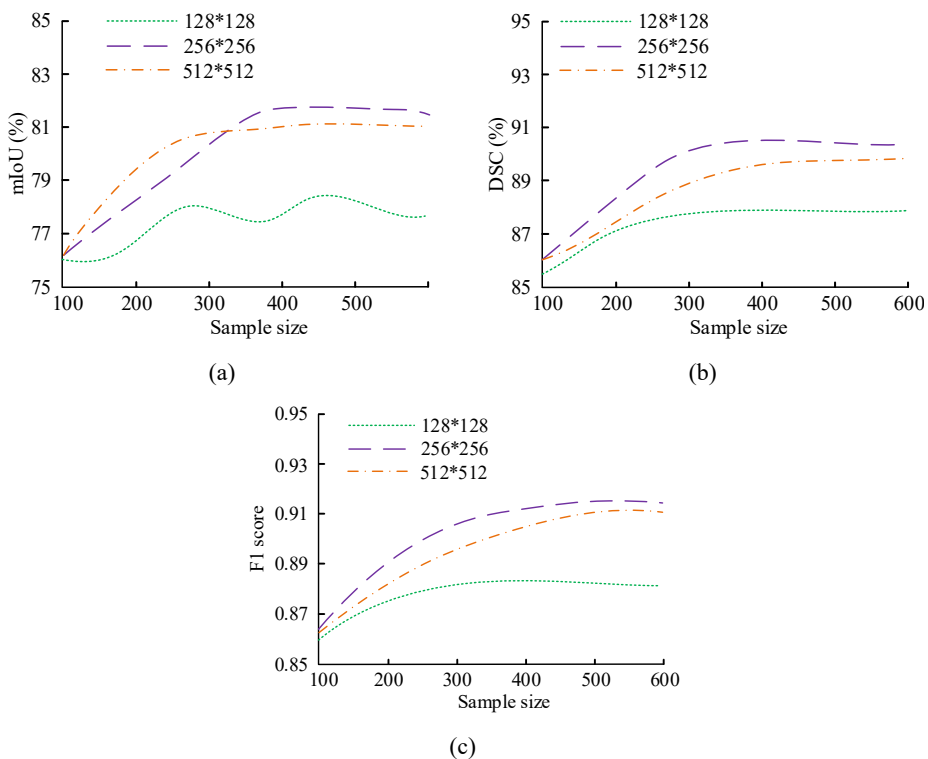


Table 1 shows that removing context modelling reduced the mIoU to 77.4%, DSC to 87.0%, and F1 to 0.86, indicating its foundational role in semantic relevance and global

structural coherence. Removing the multi-scale attention module slightly increased the metrics (mIoU 78.6%, DSC and F1 marginally improved), showing that although multi-scale fusion aids texture perception, single-scale features still limit detail consistency. When the directional enhancement mechanism was removed, mIoU rose to 79.5%, DSC to 89.1%, and F1 to 0.89, yet all remained below the full model, confirming the necessity of orientation-aware modelling for brushstroke and edge reconstruction in heritage imagery. Combined ablations (e.g., mIoU dropping to 75.6% when context modelling and orientation enhancement are both removed) resulted in more pronounced degradation, demonstrating that the three modules functioned synergistically rather than independently. The complete CA-MSI-Net achieved the best performance – mIoU 81.5%, DSC 90.8%, F1 0.91 – verifying that context understanding, multi-scale fusion, and directional enhancement jointly strengthened semantic expression and detail restoration. The stability test results of the model at different resolutions are shown in Figure 10.

As shown in Figure 10(a), the overall performance of models with input resolutions of 256×256 and 512×512 on the mIoU metric was better than that of 128×128 . Among them, 256×256 reached its peak (about 81.5%) after the sample size reached 400, slightly higher than 512×512 (about 81.3%), while 128×128 always maintained between 77% and 79%, with relatively low performance. From Figure 10(b), in terms of DSC indicators, the 256×256 resolution maintained the highest value throughout the entire process, reaching up to 90.8%, followed by 512×512 (about 90.6%), and 128×128 was significantly lower (about 88.4%). According to Figure 10(c), both 256×256 and 512×512 were significantly better than 128×128 in the F1-score index, and tended to stabilise after the sample size exceeds 300. Among them, 256×256 was slightly higher (91.2%) than 512×512 (91.0%). The experimental results showed that medium resolution (256×256) could balance computational efficiency and accuracy performance in the model, and was suitable for most reconstruction scenarios, while high-resolution had higher computational costs with limited performance improvement. The comparison results of inference speed and perceptual quality of different models on various hardware platforms are shown in Table 2.

Table 2 Inference speed and perceptual quality comparison of different models on various hardware platforms

<i>Model</i>	<i>A100 GPU inference time (ms)</i>	<i>RTX 3080 Ti inference time (ms)</i>	<i>Jetson AGX Orin inference time (ms)</i>	<i>LPIPS</i>
U-Net3+	12.5	18.3	46.7	0.142
Swin-transformer U-Net	16.8	24.5	61.2	0.125
UCTransNet	14.9	21.6	57.3	0.131
MSAU-Net	11.2	16.5	42.8	0.138
DovNet	12.0	17.9	44.1	0.136
CA-MSI-Net	11.8	16.9	41.5	0.118

As shown in Table 2, CA-MSI-Net demonstrated clear advantages in both inference efficiency and perceptual quality. On the A100 platform, its average inference time was 11.8 ms – close to MSAU-Net (11.2 ms) and faster than UCTransNet (14.9 ms) and Swin-transformer U-Net (16.8 ms) – indicating efficient computation with strong

structural modelling. On the RTX 3080 Ti, the inference time was 16.9 ms, about 3.0–7.6 ms faster than transformer-based models. On the Jetson AGX Orin, CA-MSI-Net reached 41.5 ms, outperforming U-Net3+ and UCTransNet by 5.2 ms and 15.8 ms, showing good adaptability for lightweight deployment. In perceptual quality evaluation, CA-MSI-Net achieved the lowest LPIPS (0.118), outperforming U-Net3+ (0.142) and DovNet (0.136), reflecting stronger texture fidelity and visual consistency. Although MSAU-Net was slightly faster, its LPIPS (0.138) was noticeably worse, revealing limitations in reconstruction quality. Swin-transformer U-Net exhibited better global modelling but had the slowest inference speed across all platforms, reducing its suitability for time-critical reconstruction. Overall, CA-MSI-Net offered the best perceptual performance while maintaining a balanced inference speed and resource usage, confirming its advantages in visual quality, real-time responsiveness, and engineering deployment for digital reconstruction of intangible cultural heritage paintings.

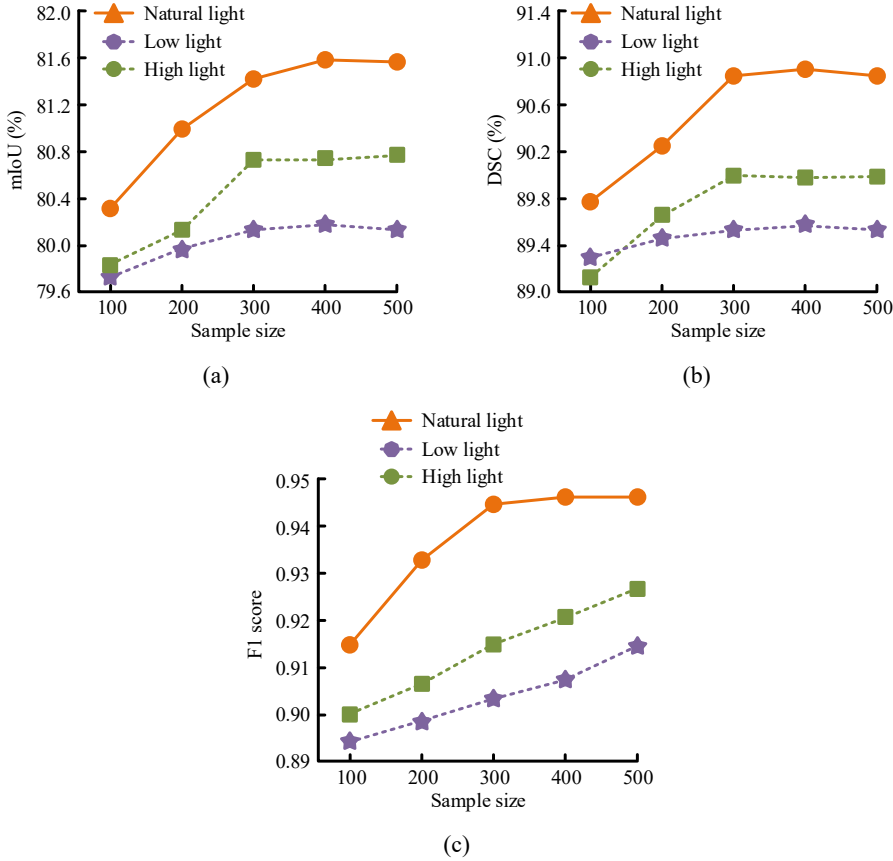
3.2 Application effect analysis of MSI image reconstruction method for intangible cultural heritage paintings based on CA-MSI-Net

To comprehensively verify the performance of the CA-MSI-Net model in the MSI image reconstruction task of intangible cultural heritage paintings, experimental analysis was conducted from four aspects: overall reconstruction accuracy evaluation, comparison of model performance, analysis of reconstruction effects in different bands, and robustness testing in high noise scenarios. The experimental data were all sourced from the standardised preprocessed MSI dataset of intangible cultural heritage paintings, with evaluation metrics including mIoU, DSC, F1-score, root mean square error (RMSE), and peak signal to noise ratio (PSNR). All experiments were conducted under the same hardware environment and training parameters. The evaluation results of reconstruction stability under different lighting conditions are shown in Figure 11. To better simulate real acquisition conditions, lighting configurations were designed according to scenarios commonly encountered in cultural heritage digitisation. Natural-light settings represent standard non-contact acquisition environments in open galleries or daylight-assisted restoration studios, where diffuse illumination is available. High-light settings ($\approx 1,200$ – $1,500$ lux) emulate professional studio lighting used for fine pigment examination and high-resolution documentation. Low-light settings (below 300 lux) reproduce on-site acquisition conditions in which strong illumination must be avoided to prevent photodegradation, resulting in reduced ambient brightness.

As shown in Figure 11(a), the model performed best under natural light conditions in terms of mIoU index. With the increase of sample size, mIoU gradually increased from about 80.6% to 81.6%, and tended to stabilise after the sample size reached 400. Under high light conditions, mIoU was generally lower than natural light and ultimately stabilised at about 81.0%. Under low light conditions, the performance was the worst, fluctuating only between 79.6% and 80.2%, indicating that low light significantly affected the accuracy of structural restoration. From Figure 11(b), in terms of DSC indicators, natural light also maintained the highest level, with a peak of about 91.0%, followed by high light conditions (about 90.2%), and low light conditions (about 89.6%), indicating that changes in light intensity could affect boundary consistency. From Figure 11(c), in the F1-score index, the model reached a maximum of about 0.95 under natural light conditions, and remained stable after the sample size exceeded 300. Under high

light conditions, it gradually increased to about 0.92. The amplification was limited under low light conditions, with a maximum of about 0.91. Experimental data showed that CA-MSI-Net could achieve the best reconstruction effect under natural light, followed by high light. Low light conditions significantly suppressed the model's performance in structural restoration, boundary consistency, and detail recovery. The evaluation results of the model in different application scenarios are shown in Figure 12.

Figure 11 Evaluation of reconstruction stability under different lighting conditions, (a) MIoU under different lighting conditions (b) DSC under different lighting conditions (c) F1-score under different lighting conditions (see online version for colours)



According to Figure 12(a), in terms of PSNR index, the colour consistency of the silk fading area reconstruction scene always maintained the highest level, and PSNR remained stable above 35 dB, with small fluctuations with sample size changes, indicating that the model had high colour restoration and detail preservation capabilities in this type of scene. The PSNR of both the restoration of details in mural painting and the completion of local stain information remained between 34 dB and 34.5 dB, with stable performance but slightly lower than the fading area of silk. The PSNR of Gongbi high-density texture and microcrack repair scenes was the lowest (about 33.8–34.2 dB), and the fluctuation amplitude was relatively larger when the sample size increased. According to Figure 12(b), in the RMSE index, the reconstruction of the silk fading area

also performed well, with a stable RMSE of around 0.019 and a small error. The RMSE for the restoration of details and completion of local stain information in the mural remained between approximately 0.020 and 0.021, respectively. The RMSE of Gongbi high-density texture and microcrack repair was the highest (about 0.023–0.024), and the fluctuation range was relatively large. Experimental data showed that CA-MSI-Net had the most significant advantage in colour consistency reconstruction of silk fading areas, maintaining high PSNR and low RMSE at different sample sizes. The comparison results of reconstruction performance under different styles of intangible cultural heritage painting are shown in Table 3.

Figure 12 Evaluation results of the model in different application scenarios (see online version for colours)

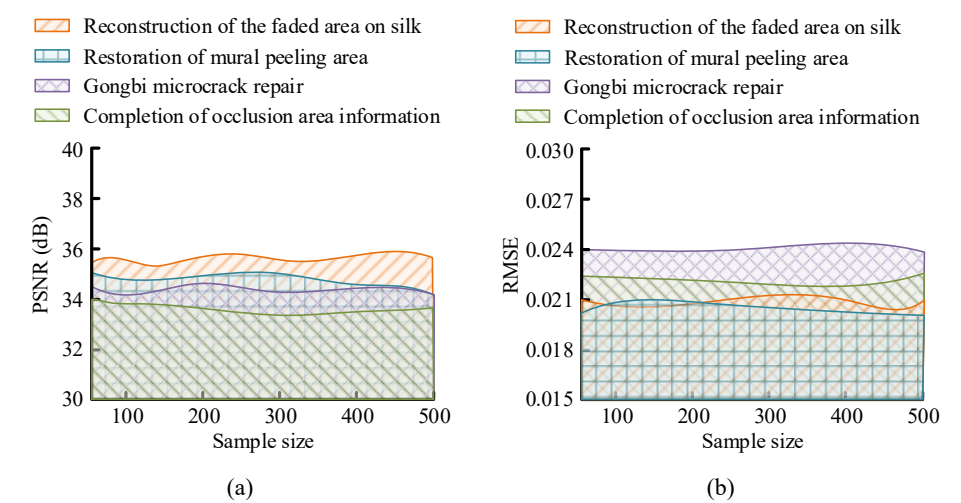


Table 3 Comparison of reconstruction performance under different intangible cultural heritage painting styles

Painting style	Model	mIoU (%)	DSC (%)	F1-score
Gongbi	CA-MSI-Net	82.4	91.6	0.92
	UCTransNet	78.3	88.5	0.89
	DovNet	79.1	89.2	0.90
Mural	CA-MSI-Net	80.9	90.2	0.91
	UCTransNet	77.1	87.9	0.88
	DovNet	78.0	88.6	0.89
Silk painting	CA-MSI-Net	81.6	90.8	0.91
	UCTransNet	77.5	88.0	0.88
	DovNet	78.4	88.8	0.89

According to Table 3, among the three typical intangible cultural heritage painting styles, CA-MSI-Net significantly outperformed the comparative model in all indicators. In the meticulous style, the mIoU, DSC, and F1-scores of CA-MSI-Net were 82.4%, 91.6%, and 0.92, respectively, which were 4.1%, 3.1%, and 0.03% higher than UCTransNet, and

3.3%, 2.4%, and 0.02% higher than DovNet, respectively. In the mural style, the mIoU, DSC, and F1-scores of CA-MSI-Net were 80.9%, 90.2%, and 0.91, respectively, which were 3.8%, 2.3%, and 0.03 higher than UCTransNet and 2.9%, 1.6%, and 0.02 higher than DovNet. Under the silk style, the three indicators of CA-MSI-Net were 81.6%, 90.8%, and 0.91, respectively, which also performed the best. Experimental data showed that CA-MSI-Net exhibited stable and significant performance advantages in different styles of intangible cultural heritage painting, verifying its adaptability and robustness in multiple types of scenes. The comparison of reconstruction performance under different spectral channel numbers is shown in Table 4.

Table 4 Comparison of reconstruction performance under different spectral channel numbers

<i>Spectral channels</i>	<i>mIoU (%)</i>	<i>DSC (%)</i>	<i>F1-score</i>	<i>FLOPs (G)</i>
6 channels	79.2	88.9	0.89	65.4
8 channels	80.4	89.8	0.90	70.1
12 channels	81.5	90.8	0.91	76.5
16 channels	81.6	90.9	0.91	84.3

According to Table 4, there were certain differences in the reconstruction performance of CA-MSI-Net under different spectral channel numbers. The mIoU, DSC, and F1-score all showed a steady improvement trend as the number of channels increased from 6 to 12. The performance was best at 12 channels, with mIoU at 81.5%, DSC at 90.8%, and F1-score at 0.91, indicating that the configuration achieved the best balance between structural restoration and spectral consistency. At 16 channels, the three accuracy indicators were almost the same as those at 12 channels (mIoU increased by 0.1%, DSC and F1-score did not show significant changes), but FLOPs significantly increased to 84.3 G, and the computational cost increased by about 9.7% compared to 12 channels, indicating that performance benefits tended to saturate at high channel numbers. The accuracy of channels 6 and 8 was significantly lower than that of channels 12, especially for channels 6 where mIoU was only 79.2%, DSC was 88.9%, and F1-score was 0.89. However, their FLOPs were the lowest, only 65.4 G, making them suitable for scenarios that require high computing resources. Experimental data showed that 12 channels were the best compromise between performance and efficiency for the current task, ensuring high-quality reconstruction while controlling computational overhead.

4 Discussion

In response to the challenges of spectral feature fidelity and multi-style detail restoration in the digital reconstruction of intangible cultural heritage paintings, the CA-MSI-Net reconstruction framework demonstrated significant advantages in multispectral feature extraction, cross-channel information fusion, and detail preservation. The experimental results revealed that the proposed model outperformed comparative methods in terms of average PSNR, SSIM, and spectral consistency index across different intangible cultural heritage painting styles. The PSNR remained stable above 32 dB, while the spectral consistency index improved by over 5%. This demonstrated that CA-MSI-Net achieved high stability and robustness in cross-style, cross-condition scenarios, overcoming the limitations of traditional models, which exhibited significant performance fluctuations in

complex texture restoration and under varying lighting conditions. Compared with existing approaches, the multispectral hierarchical convolution network of Chen et al. (2024) offered improved colour accuracy but relied on fixed receptive fields and lacked long-range dependency modelling. Consequently, complex structures and multilayer pigment deposits often cause blurred details, colour drift, and weak cross-style adaptability, especially in high-texture scenes. Xin et al.'s (2025) method combined spectral mapping with attention and performed reasonably on multi-style datasets, yet its attention remained local and failed to capture global structural relations. Its purely end-to-end training strategy also lacked hierarchical feature regulation, leading to slower inference on high-resolution inputs and limited fine-texture recovery under challenging illumination.

In contrast, CA-MSI-Net introduced a spectral-spatial co-modelling mechanism. A spatial Transformer enhanced long-range structural inference, while a channel transformer strengthened cross-band coupling, effectively improving spectral consistency and texture hierarchy reconstruction. The model further adopted progressive multi-scale guidance and contextual enhancement, enabling stronger adaptation to diverse styles, texture densities, and illumination conditions. Experiments showed that CA-MSI-Net maintained stable performance across meticulous brushwork, silk paintings, and mural datasets, improving PSNR and spectral consistency by over 5% compared with other methods. Moreover, it achieved lower FLOPs and faster unit inference time than most transformer-based models, demonstrating both high reconstruction accuracy and strong deployment efficiency.

Additionally, experimental results showed that CA-MSI-Net achieved an optimal balance between performance and efficiency under a 12-channel spectral configuration, with significant improvements in mIoU and DSC while outperforming most comparative models in FLOPs. These findings indicated that the proposed method could ensure high precision while managing computational costs, making it feasible for practical deployment in museum digital archiving, online exhibitions, and mobile terminal interactions. Under natural lighting conditions, the model demonstrated superior adaptability, maintaining high levels of edge detail and colour consistency in its reconstruction results, thereby showcasing its feasibility and potential for widespread application in real-world acquisition environments. Overall, CA-MSI-Net not only complemented the performance of existing digital reconstruction methods for intangible cultural heritage paintings but also provided a novel pathway for the digital preservation, dissemination, and reuse of cultural heritage. Beyond high-precision reconstruction of intangible cultural heritage paintings, CA-MSI-Net also showed broad applicability in other digital preservation tasks. Textiles, crafts, murals, and weathered surfaces often contain complex textures and subtle colour gradients, and are easily affected by lighting and environmental noise – making traditional RGB or single-depth models insufficient for high-fidelity restoration. The multispectral feature decomposition and spectrum-space co-modelling in CA-MSI-Net enabled effective cross-domain reconstruction of pigment layers, material aging patterns, and structural cracks. With transfer learning or domain-adaptation strategies, the method can further extend to grotto mural restoration, repair of corroded manuscript areas, and other archival or virtual-display applications, supporting digital documentation and post-disaster assessment. Overall, CA-MSI-Net offers strong scalability and has the potential for wider adoption in cultural heritage protection systems.

5 Conclusions

The digital reconstruction of intangible cultural heritage paintings requires a balance between detail fidelity and spectral consistency under various complex conditions, which traditional RGB and a single deep network cannot meet. The study integrated an MSI system with an improved U-Net architecture to propose the CA-MSI-Net reconstruction method. This approach enhanced spatial and spectral feature representation through the collaborative use of spatial transformer and inter-spectral transformer modules, while incorporating contextual modelling and multiscale attention mechanisms to improve structural perception and detail restoration capabilities. Unlike traditional approaches that focus solely on spatial reconstruction or spectral calibration, CA-MSI-Net achieves stronger performance through a unified spectral-spatial co-modelling strategy and multi-scale context-aware design. The spatial transformer expands the effective receptive field to reinforce global structural consistency, while the channel transformer models long-range spectral dependencies to better represent pigment-layer variations and subtle colour transitions. In addition, the multi-scale attention mechanism and directional context modelling jointly enhance hierarchical texture recovery and structural continuity, enabling accurate reconstruction of brushstroke patterns and directional visual features characteristic of intangible cultural heritage paintings. Experimental results demonstrated that, in terms of overall reconstruction accuracy, CA-MSI-Net achieved mIoU, DSC, and F1-score values of 81.3%, 92.1%, and 0.94, respectively, outperforming comparative methods such as U-Net3+, UCTransNet, and DovNet. Regarding adaptability to different painting styles, the model exhibited leading performance in meticulous brushwork (mIoU 82.4%, DSC 91.6%, F1-score 0.92), murals (80.9%, 90.2%, 0.91), and silk-mounted paintings (81.6%, 90.8%, 0.91). In spectral channel experiments, the 12-channel configuration achieved an optimal balance (mIoU 81.5%, DSC 90.8%, F1-score 0.91, FLOPs 76.5 G). Under lighting condition tests, mIoU improved to 81.6% and F1-score reached 0.95 under natural light. Although CA-MSI-Net demonstrated exceptional performance in multiscenario reconstruction, its performance declined under extreme low-light and high-noise conditions. Additionally, while higher channel counts yielded some accuracy improvements, they also resulted in significant increases in computational costs. In the future, optimisation will be carried out for low light enhancement, high noise suppression, and cross resolution modelling, and lightweight transformer structures and adaptive channel selection mechanisms will be explored to further reduce computational costs while ensuring reconstruction accuracy, expanding its real-time application potential in large-scale cultural heritage protection.

Declarations

The author confirms that there are no relevant financial or non-financial competing interests to report.

References

- Zhang, F. (2023) 'Quantitative analysis of color characteristics for Yangjiabu New Year woodblock paintings based on color science and spectral imaging', *Color Res. Appl.*, September, Vol. 48, No. 6, pp.801–813, DOI: 10.1002/col.22896.
- Guo, J. (2023) 'Inheritance, cultural legacy and innovation in traditional art: a case study of Chinese innerpainted snuff bottles', *Glass Technol.-Eur. J. Glass Sci. Technol. Part A*, August, Vol. 64, No. 4, pp.120–132, DOI: 10.13036/17533546.64.4.008.
- Liu, K., Wu, H., Gao, Y., Zhu, C., Ji, Y. and Lü, Z. (2023) 'Archaeology and virtual simulation restoration of costumes in the Han Xizai banquet painting', *Autex Res. J.*, March, Vol. 23, No. 2, pp.238–252, DOI: 10.2478/aut-2022-0001.
- Zhang, X., Cui, G., Ruan, X., Cui, D., Gao, X. and Chen, Q. (2024a) 'Spectral reflectance reconstruction based on wideband multi-illuminant imaging and a modified particle swarm optimization algorithm', *Opt. Express*, January, Vol. 32, No. 3, pp.2942–2958, DOI: 10.1364/OE.506136.
- Rathi, V., Rana, K. and Goyal, P. (2024) 'Iterative spectral correlation based multispectral image demosaicking', *Signal, Image and Video Process.*, July, Vol. 18, No. 11, pp.7873–7886, DOI: 10.1007/s11760-024-03435-3.
- Ma, Z. and Ng, M.K. (2025) 'Multispectral image restoration by generalized opponent transformation total variation', *SIAM J. Imaging Sci.*, July, Vol. 18, No. 1, pp.246–279, DOI: 10.1137/24M1647758.
- Rihani, N. (2023) 'Interactive immersive experience: digital technologies for reconstruction and experiencing temple of Bel using crowdsourced images and 3D photogrammetric processes', *Int. J. Archit. Comput.*, April, Vol. 21, No. 4, pp.730–756, DOI: 10.1177/14780771231168224.
- Belhi, A., Bouras, A., Al-Ali, A.K. and Fofou, S. (2023) 'A machine learning framework for enhancing digital experiences in cultural heritage', *J. Enterprise Inf. Manage.*, May, Vol. 36, No. 3, pp.734–746, DOI: 10.1108/JEIM-02-2020-0059.
- Xiang, C., Yang, Y., Zhou, T. and Wang, T. (2024) 'Digital reconstruction of historical cultural landscapes based on image recognition technology', *Traitement Signal*, June, Vol. 41, No. 3, pp.889–897, DOI: 10.18280/ts.410336.
- Zhang, W. and Taib, N. (2025) 'Virtual experimental model and digital preservation path of traditional wooden carvings in Huizhou traditional dwellings', *Int. J. Conserv. Sci.*, January, Vol. 16, No. 1, pp.199–206, DOI: 10.36868/IJCS.2025.01.13.
- Zhang, Z. and Cao, Y. (2025) 'Dunhuang Faces: a dataset enhanced through image processing for cultural heritage and machine learning', *Traitement du Signal*, February, Vol. 42, No. 1, pp.569–581, DOI: 10.18280/ts.420148.
- Al Saad, Z. (2024) '3D documentation and visualization of Greco-Roman monuments from Jordan through the application of digital technologies', *Egypt. J. Archaeol. Rest. Stud.*, December, Vol. 14, No. 2, pp.217–221, DOI: 10.21608/ejars.2024.396688.
- Gao, Y., Liu, Q., Gu, L. and Fu, Y. (2025) 'Grayscale-assisted RGB image conversion from near-infrared images', *Tsinghua Sci. Tech.*, October, Vol. 30, No. 5, pp.2215–2226, DOI: 10.26599/TST.2024.9010115.
- Mu, Q., Wang, X., Wei, Y. and Li, Z. (2021) 'Low and non-uniform illumination color image enhancement using weighted guided image filtering', *Comput. Vis. Media*, December, Vol. 7, No. 4, pp.529–546, DOI: 10.1007/s41095-021-0232-x.
- Lei, Y., Zhang, Q., Guo, Y., Pu, M., Zou, F. and Li, X. (2023) 'Snapshot multi-dimensional computational imaging through a liquid crystal diffuser', *Phot. Res.*, March, Vol. 11, No. 3, pp.B111–B124, DOI: 10.1364/PRJ.476317.
- Chen, Y., Xing, F., Li, W., Sheng, Y., Di, Y. and Gan, Z. (2022) 'Snapshot multi-frame parallel spectral holographic microscopy based on a reconfigurable optical comb', *Opt. Lett.*, December, Vol. 47, No. 24, pp.6468–6471, DOI: 10.1364/OL.479993.

- Zhang, X., Gao, Y., Ji, Y., Feng, A. and Zhao, S. (2024b) ‘Compact multi-spectral-resolution Wynne-Offner imaging spectrometer with a long slit’, *Appl. Opt.*, February, Vol. 63, No. 6, pp.1577–1582, DOI: 10.1364/AO.517374.
- Shafique, A., Seydi, S.T. and Cao, G. (2023) ‘BCD-Net: building change detection based on fully scale connected U-Net and subpixel convolution’, *Int. J. Remote Sens.*, December, Vol. 44, No. 23, pp.7416–7438, DOI: 10.1080/01431161.2023.2285737.
- Lee, Y., Kim, S.H., Noh, Y.J. and Kim, J.H. (2023) ‘Deep learning-based summertime turbulence intensity estimation using satellite observations’, *J. Atmos. Oceanic Technol.*, November, Vol. 40, No. 11, pp.1433–1448, DOI: 10.1175/JTECH-D-22-0137.1.
- Duarte, K.T.N., de Barros, M.C., Sidhu, A.S., Gobbi, D., McCreary, C. and Saad, F. (2025) ‘Attention blocks improve white matter hyperintensity semantic segmentation using U-Nets’, *IEEE Lat. Am. Trans.*, August, Vol. 23, No. 8, pp.646–661, DOI: 10.1109/TLA.2025.11072499.
- Park, J., Kim, S., Seol, S.J. and Byun, J. (2025) ‘Improving generalization performance of deep learning-based seismic data interpolation’, *Geophy. Prospect.*, September, Vol. 73, No. 5, pp.1534–1551, DOI: 10.1111/1365-2478.70020.
- Li, P., Wang, Y., Si, T., Ullah, K. and Han, W. (2023) ‘DSFA: cross-scene domain style and feature adaptation for landslide detection from high spatial resolution images’, *Int. J. Digit. Earth*, January, Vol. 16, No. 1, pp.2426–2447, DOI: 10.1080/17538947.2023.2229794.
- Wu, W., Liu, G., Liang, K. and Zhou, H. (2023) ‘Inner cascaded U²-Net: an improvement to plain cascaded U-Net’, *CMES – Comput. Model. Eng. Sci.*, June, Vol. 134, No. 2, pp.433–452, DOI: 10.32604/cmes.2022.020428.
- Valdez, M.A., Rezac, J.D., Wakin, M.B. and Gordon, J.A. (2024) ‘Multi-frequency spherical near-field antenna measurements using compressive sensing’, *IEEE J. Sel. Top. Signal Process.*, July, Vol. 18, No. 4, pp.572–586, DOI: 10.1109/JSTSP.2024.3424310.
- Zhao, H., Zhang, L., Zhang, Y. and Wang, Y. (2024) ‘Imagery overlap block compressive sensing with convex optimization’, *IEEE Trans. Intell. Transp. Syst.*, July, Vol. 25, No. 7, pp.8076–8092, DOI: 10.1109/TITS.2024.3376455.
- Nouasria, H. and Et-tolba, M. (2022) ‘A fast gradient-based sensing matrix optimization approach for compressive sensing’, *Signal Image Video Process.*, December, Vol. 16, No. 8, pp.2279–2286, DOI: 10.1007/s11760-022-02193-4.
- Banu, A.S. and Deivalakshmi, S. (2023) ‘AWUNet: leaf area segmentation based on attention gate and wavelet pooling mechanism’, *Signal Image Video Process.*, May, Vol. 17, No. 5, pp.1915–1924, DOI: 10.1007/s11760-022-02403-z.
- Pal, S., Roy, A., Palaiahnakote, S. and Pal, U. (2023) ‘Adapting a swin transformer for license plate number and text detection in drone images’, *Artificial Intelligence and Applications*, September, Vol. 1, No. 3, pp.129–138, DOI: 10.47852/bonviewAIA3202549.
- Chen, P., Wang, H., Su, X., Yang, J. and Wei, C. (2024) ‘Pseudo-panchromatic image guided transformer model for multispectral image demosaicing’, *Opt. Appl.*, April, Vol. 54, No. 4, pp.345–356, DOI: 10.37190/oa/193937.
- Xin, L., Ma, X., Luo, H., Hu, X., Xiao, K. and Liang, J. (2025) ‘Low-light image enhancement via multispectral reconstruction’, *J. Imaging Sci. Technol.*, March, Vol. 69, No. 2, pp.123–134, DOI: 10.2352/J.ImagingSci.Technol.2025.69.2.020503.