



A Comprehensive Method for Art Style Analysis and Automatic Classification Based on Multi-Channel Feature Extraction and Fusion

Jie Gao¹, Liangyi Li^{2*}, Kaiyue Xing³

¹ College of Arts, Cheongju University, Cheongju 28503, Republic of Korea

² Fine and Applied Arts, Bangkok Thonburi University, Bangkok 10170, Thailand

³ School of Computer Science and Engineering, Changchun University of Technology, Changchun 130012, China

Corresponding Author Email: liliangyi025@163.com

Copyright: ©2025 The authors. This article is published by IETA and is licensed under the CC BY 4.0 license (<http://creativecommons.org/licenses/by/4.0/>).

<https://doi.org/10.18280/ts.420503>

Received: 8 January 2025

Revised: 20 June 2025

Accepted: 18 July 2025

Available online: 31 October 2025

Keywords:

art style analysis, automatic classification, multi-channel feature extraction, feature fusion, model lightweight design

ABSTRACT

The rapid advancement of digital technologies has led to an exponential increase in image data, spanning various domains such as painting, illustration, animation, and photography. Efficient and accurate analysis and classification of art styles are critically needed in contexts such as art management, digital media content retrieval, and copyright protection. Traditional manual classification methods, however, are time-consuming, labor-intensive, and prone to subjectivity, rendering them inadequate for large-scale and high-precision applications. As a result, automatic art style analysis and classification have become key challenges within the fields of computer vision and image processing. Existing methods for art style analysis and classification exhibit several limitations: some rely solely on a single feature, resulting in an incomplete description of the art style and reduced classification accuracy; others involve complex model structures and excessive parameters, leading to low efficiency and difficulties in meeting real-time application demands; furthermore, many approaches struggle to adapt to complex, hybrid, or variant art styles, thereby compromising classification stability. To address these issues, a novel method for comprehensive art style analysis and automatic classification is proposed, based on multi-channel feature extraction and fusion. This method consists of three critical modules: (1) Model Lightweight Design, which simplifies the structure and reduces parameters, thus improving efficiency while maintaining accuracy, making it suitable for large-scale data processing; (2) Feature Extraction and Enhancement, where multiple channels are used to extract and enhance features related to color, texture, shape, and composition, capturing the essence of the art style comprehensively; (3) Multi-channel Feature Fusion, which effectively combines the extracted features from different channels, utilizing complementary information to enhance the recognition of complex art styles. The innovations of this method are as follows: first, multi-channel feature extraction and enhancement overcome the limitations of single-feature approaches, enabling a more holistic description of art styles; second, the lightweight model design ensures both efficiency and accuracy, addressing the bottlenecks of traditional complex models; and third, multi-channel feature fusion improves adaptability to complex, mixed, and variant art styles, thus enhancing classification stability. This research presents an efficient and precise solution for automatic art style analysis and classification, which holds the potential to advance practical applications in art research and the digital media industry.

1. INTRODUCTION

With the rapid development of digital technology, image data has been growing explosively, covering many fields such as painting, illustration, animation, and photography [1-3]. In the management of art works [4, 5], museums and galleries need to effectively organize and classify a large number of digital art collections for research and exhibition purposes. In the field of digital media content retrieval [6, 7], users hope to quickly and accurately find image resources that match specific art styles, such as designers searching for materials of a particular style or art enthusiasts looking for works in their preferred style. In copyright protection [8, 9], accurately

identifying different art styles helps determine the originality and similarity of works. However, traditional manual analysis and classification of art styles are not only time-consuming and labor-intensive but also highly influenced by subjective factors, making them difficult to meet the demands of large-scale and high-precision applications. Therefore, the automatic analysis and classification of art styles has become an important research topic in the field of computer vision and image processing.

Related research holds significant value in various fields. In art research [10], automatic analysis and classification of art styles can provide new perspectives and tools for art history studies, helping researchers more objectively organize the

development and stylistic features of different art movements and discover stylistic relationships between different artists. In the digital media industry [11], efficient art style analysis and classification technologies can enhance content production and management efficiency. For example, in animation production, it allows for the rapid selection of materials that match the project style, reducing production costs. In e-commerce platforms [12], it can precisely recommend product images that match users' aesthetic preferences, improving user experience. Furthermore, this research can promote the deeper development of artificial intelligence in the field of image understanding, enriching the ability of computers to interpret visual content.

Existing research methods in art style analysis and classification have numerous shortcomings and limitations. Some methods focus solely on the extraction of a single feature, such as color or texture, neglecting other important features such as shape and composition. This leads to an incomplete description of the art style and subsequently impacts classification accuracy, as highlighted by methods in references [13, 14]. Some models are structurally complex and have too many parameters, resulting in low efficiency and difficulty applying them in real-time scenarios. Models in references [15-17] exhibit poor performance when handling large-scale data due to their excessive complexity. Some methods also struggle to adapt to complex and evolving art styles, and the classification accuracy drops significantly when dealing with mixed or mutated art styles, as indicated in references [18-20], where traditional methods showed noticeably reduced accuracy when processing images that combine multiple styles.

This paper proposes a comprehensive art style analysis and automatic classification method based on multi-channel feature extraction and fusion, which includes three key ideas. The model lightweight design aims to simplify the model structure and reduce the number of parameters, improving the model's operational efficiency while maintaining analysis accuracy, enabling it to meet the needs of real-time processing and large-scale data applications. The feature extraction and enhancement module extracts various features from images such as color, texture, shape, and composition through multiple channels and enhances the extracted features, compensating for the shortcomings of single-feature descriptions and capturing the essential characteristics of the

art style more comprehensively. The multi-channel feature fusion module effectively combines the features extracted and enhanced from different channels, fully utilizing complementary information between features to improve the recognition and classification ability of complex art styles. The value of this research lies in the fact that the proposed method effectively solves the issues found in existing studies, such as single-feature reliance, low efficiency, and poor adaptability, improving the accuracy and efficiency of art style analysis and automatic classification. This method provides more reliable technical support for applications in related fields and promotes the practical implementation and development of art style analysis and classification technologies.

2. METHOD DESIGN

This paper focuses on the research of multi-channel feature extraction and fusion methods, primarily due to the complexity and multi-dimensional characteristics of the comprehensive art style of artworks. The style of a painting is the result of the combined effects of various visual elements, such as color tone, brushstroke texture, compositional logic, and line features. A single feature dimension is insufficient to fully capture the essence of the style. For example, the light and shadow color changes of Impressionist works and the dynamic composition of Baroque style belong to different feature dimensions, and color features alone cannot distinguish the rigor of Classicism from the simplification of Neoclassicism. Similarly, relying solely on texture features makes it difficult to differentiate the emotional brushstrokes of Expressionism from the random brushstrokes of Abstract Expressionism. Multi-channel feature extraction can independently model different style dimensions, such as capturing the color rhythm of Monet's works through the color channel, extracting the brushstroke intensity of Van Gogh through the texture channel, and analyzing Picasso's deconstruction techniques of Cubism through the shape channel. The fusion process can then integrate the scattered feature information, forming complementary characteristics and constructing a more comprehensive style description system, thus overcoming the limitations of single-feature representations of complex art styles.

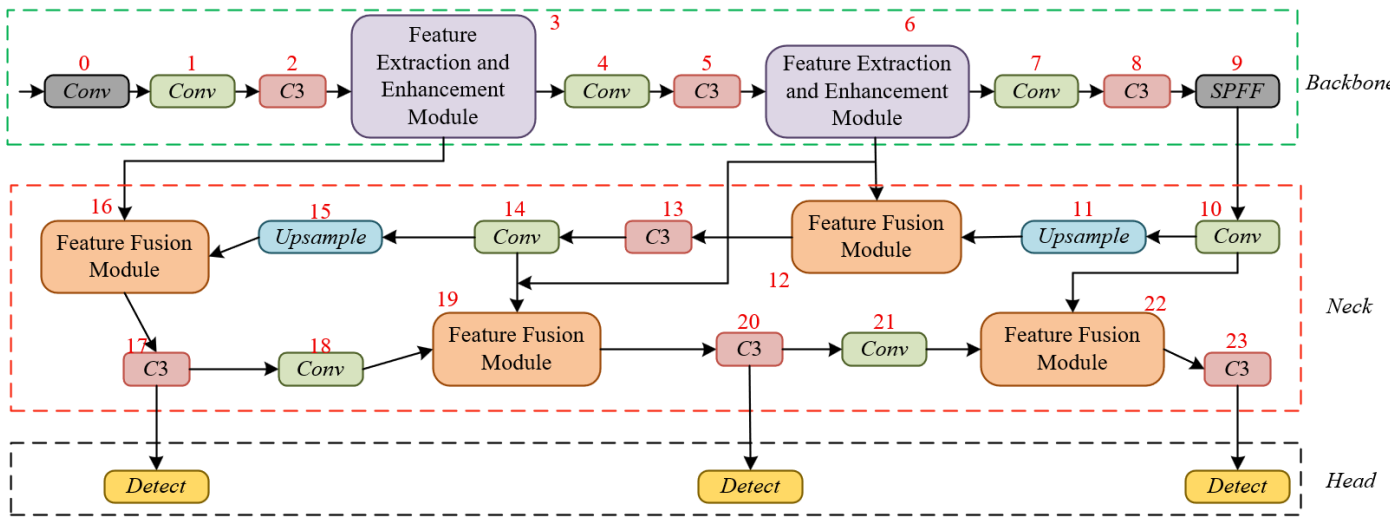


Figure 1. Structure diagram of the comprehensive art style analysis and automatic classification model based on multi-channel feature extraction and fusion

There are several challenges in the feature extraction and fusion of the comprehensive art style of artworks. First, the precision of feature extraction is difficult to achieve. The style of a painting is often highly subjective and creative, and variations in the style of different artists within the same school or cross-style fusion lead to blurred feature boundaries, which may result in feature overlap or misjudgment. Second, the adaptability of fusion strategies is a major issue. The feature weights of different style dimensions are dynamic. For instance, in realism, the weight of shape and composition is higher, while in Fauvism, the weight of color predominates. How to dynamically adjust the fusion ratio to accommodate diverse art styles and avoid feature redundancy or loss of critical information is a core problem that needs to be solved. In addition, differences in the medium used to create the artwork and its preservation state may introduce noise features, further increasing the difficulty of feature extraction and fusion.

To address these challenges, this paper designs a comprehensive art style analysis and automatic classification method based on multi-channel feature extraction and fusion. This method includes three key ideas: model lightweight design, multi-scale feature enhancement module, and multi-channel feature fusion module. The details are described as follows. The model structure diagram is shown in Figure 1.

2.1 Lightweight design

Comprehensive art style analysis of artworks often needs to be deployed in mobile devices or resource-constrained industrial platforms, where strict limitations exist on model memory, computation speed, and power consumption. For example, when a user takes a photo of an artwork in a gallery and performs real-time style classification, the model needs to complete multi-channel feature extraction and fusion within seconds. High-complexity models may experience delays, reducing user experience. Additionally, industrial-grade art resource management platforms need to process millions of artwork data, and a lightweight design can reduce server computing consumption and minimize hardware investment costs. Therefore, to meet the practical application requirements of artwork analysis scenarios, the comprehensive art style analysis and automatic classification model proposed in this paper adopts a lightweight design.

The core idea for achieving lightweight design in this paper is inspired by the modular architecture and multi-version adaptation strategy of YOLOv5, constructing a network structure that can be flexibly adjusted. The model is divided into two major modules: the basic feature extraction layer and the multi-channel fusion layer. The basic layer adopts a lightweight backbone network, reducing the number of convolutional kernels and simplifying the residual block structure to lower the parameter scale. At the same time, different complexity versions, such as v , t , and l , are provided. For example, when processing simple features like sketches, a lighter version is selected; when analyzing complex art styles with multi-texture fusion like oil paintings, a slightly deeper version is used, achieving "on-demand" allocation of computing resources. In addition, the Ghost module of GhostNet is introduced to replace some standard convolutions, which generates redundant feature maps through cheap operations while maintaining feature extraction capabilities, thus reducing parameter redundancy in the high-dimensional feature processing of artworks.

Furthermore, an efficient aggregation mechanism is embedded into the multi-channel feature fusion stage to avoid any negative impact of lightweight design on art style recognition accuracy. To address the complementarity of art style features, a lightweight attention mechanism is used. When fusing color and texture channel features, higher weight is dynamically assigned to the color features of Impressionist paintings and to the compositional features of Baroque paintings. This reduces the computational cost of irrelevant features while enhancing the representation of key style dimensions. Additionally, the depthwise separable convolution technology of MobileNet is adopted, which decomposes cross-channel convolutions in the multi-channel fusion process into depthwise and pointwise convolutions, thus reducing the computational complexity when different art style features interact. Through these strategies, the model can both maintain the ability to recognize subtle style differences, such as the color rhythm of Monet and the brushstroke intensity of Van Gogh, and meet the real-time analysis and large-scale processing efficiency requirements of mobile devices and industrial platforms, ultimately achieving a balance of "high precision - high efficiency" for comprehensive art style analysis.

2.2 Feature extraction and enhancement

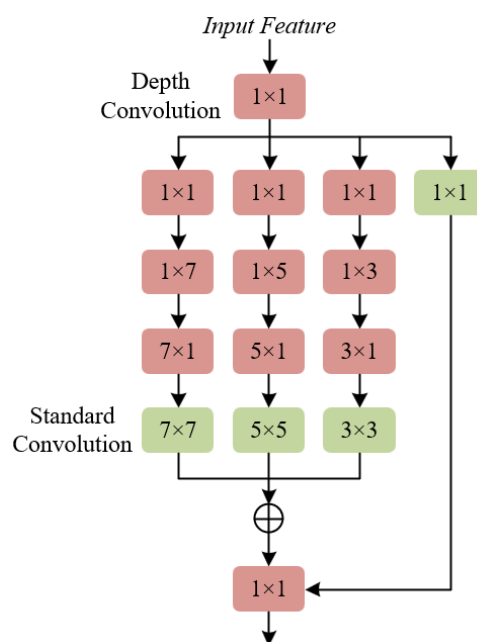


Figure 2. Structure diagram of the feature extraction and enhancement module

In the feature extraction and enhancement module, a multi-scale and multi-branch feature enhancement strategy is adopted to extract style information, targeting the multi-dimensional characteristics of the comprehensive art style of artworks. Figure 2 shows the structure diagram of the feature extraction and enhancement module. By initializing multiple branches and convolutional layers, style features are captured from different perspectives, such as brushstroke thickness, color block size, and compositional hierarchy. The channel dimensions of small-size sampling blocks are mapped to 1, and this dimensionality reduction operation effectively filters out redundant channel information unrelated to style, such as background noise and canvas texture, thereby accurately

mapping foreground style information, such as key brushstroke features and core color regions. Assuming the convolution operation of a 1×1 convolution kernel is represented by $CONV_{1 \times 1}(d)$, and the feature map for extracting preliminary information is represented by Q , the operation is as follows:

$$Q = CONV_{1 \times 1}(d) \quad (1)$$

Next, four feature extraction branches are designed, with multiple branches and convolutional layers initialized to analyze the artwork style from different dimensions. Branch 1 uses 1×1 depthwise convolution to extract local detail information of the style. This convolution method reduces the computation and memory access of redundant pixels in the artwork and more efficiently captures spatial-style features, such as point distribution in Pointillism or the brushstroke intensity variations in Expressionism. At the same time, this branch generates an equivalence feature map that retains the key style semantic information and target features, ensuring that local feature extraction does not lose core style clues. Assuming the convolution operation of the 1×1 convolution kernel is represented by $DW_CONV_{1 \times 1}$ and the output feature map of branch 1 is represented by A_1 , the operation is defined as follows:

$$A_1 = DW_CONV_{1 \times 1}(Q) \quad (2)$$

Branches 2, 3, and 4 first use 1×1 convolution to adjust the channel numbers, ensuring that the channel count of the artwork style features processed by each branch remains consistent, providing a unified foundation for subsequent cross-branch feature fusion. Branch 2 combines multi-scale convolution and depthwise convolution to capture and enhance style features from different spatial directions, such as horizontal, vertical, and diagonal, for example, analyzing the compositional blank space in landscape paintings or the color block stacking in oil paintings through multi-dimensional analysis. This ensures that the extracted feature map retains more contextual information of the style, enhancing the completeness of the style description. Assuming the output feature maps of branches 2, 3, and 4 are represented by A_2 , A_3 , and A_4 , the operation is defined as follows:

$$A_2 = \left(DW_CONV_{3 \times 3} \left(CONV_{3 \times 1} \left(CONV_{1 \times 3} \left(CONV_{1 \times 1}(Q) \right) \right) \right) \right) \quad (3)$$

Branch 3 and 4 have a similar feature extraction logic to branch 2, but use different sizes of convolution kernels to adapt to style elements at different scales in the artwork. For example, branch 3 uses a larger convolution kernel to capture the overall color blending effects in Impressionist works, while branch 4 uses a smaller convolution kernel to extract detailed line features in Realism. The differences in the convolution kernels further enrich the extracted style feature dimensions, covering the full-scale style information from macro composition to micro brushstrokes, thus enhancing the recognition ability of diverse art styles. The operations are defined as follows:

$$A_3 = \left(DW_CONV_{5 \times 5} \left(CONV_{5 \times 1} \left(CONV_{1 \times 5} \left(CONV_{1 \times 1}(Q) \right) \right) \right) \right) \quad (4)$$

$$A_4 = \left(DW_CONV_{7 \times 7} \left(CONV_{7 \times 1} \left(CONV_{1 \times 7} \left(CONV_{1 \times 1}(Q) \right) \right) \right) \right) \quad (5)$$

The style features from the outputs of the three branches are concatenated and further integrated into a unified feature representation via 1×1 convolution. This process encodes the spatial style feature information of the artwork and compresses the channel dimensions, meaning the color, brushstroke, and composition features captured by different branches are associated, and multi-dimensional style information is condensed into a compact and discriminative feature vector. This facilitates cross-channel style comparison and classification by the subsequent fusion module. The operation is defined as follows:

$$\begin{aligned} A_{CA} &= CONCAT[a_2, a_3, a_4] \\ A_{FU} &= CONV_{1 \times 1}(A_{CA}) \\ B &= RELU(\beta \cdot A_{FU} + A_1) \end{aligned} \quad (6)$$

Next, the scale features from the three branches are concatenated along the channel dimension, forming a robust feature representation that covers the multi-layered style of the artwork. Through a shortcut connection, the original style features are added to the fused feature map, followed by the application of the ReLU activation function. This not only preserves the key style information initially extracted, such as the signature color tone, but also strengthens the style differences through the fused features. Here, A_{CA} represents the feature concatenation operation, A_{FU} represents the fused style feature map, and β as a scaling factor avoids model instability caused by excessive fluctuation in style features during training, ensuring more stable feature learning for complex art styles.

2.3 Multi-channel feature fusion

In the multi-channel feature fusion module, high-level feature maps and low-level feature maps carry different dimensional information about the comprehensive art style of the artwork. High-level feature maps focus on the overall style attributes of the artwork, such as the light and shadow atmosphere of Impressionism and the symmetrical composition of Classicism, representing macro features; low-level feature maps focus on local details, such as the swirling brushstrokes in Van Gogh's works or the delicate lines in fine-line paintings, representing micro elements. By fusing the outputs of these two types of feature maps, the semantic expression ability of small-scale style elements in the artwork can be effectively enhanced. For example, when analyzing Romanticism artworks that blend delicate brushstrokes and grand compositions, the fusion operation can relate the agility of the local brushstrokes with the emotional tension of the overall composition, avoiding style information fragmentation caused by relying solely on a single feature layer.

This module is used to fuse multiple channels of the artwork feature maps, adopting the Squeeze-and-Excitation (SE) module mechanism to dynamically weight the channels of each input feature map. Figure 3 shows the structure of the SE module. For artworks of different styles, the importance of each channel feature varies significantly. For example, when analyzing Fauvist works, the high saturation features in the color channel are crucial; whereas, when analyzing Cubism

works, the shape deconstruction features in the corresponding channel are more important. The weighting mechanism enhances the expression of important feature channels while suppressing interference from irrelevant channels, such as highlighting the brushstroke intensity channel of Expressionism and diminishing the redundant background texture channel in Realism. Ultimately, the weighted feature maps are concatenated along the channel dimension, forming a fused feature that combines multi-dimensional integrity with the prominence of key features, providing a more accurate basis for style classification.

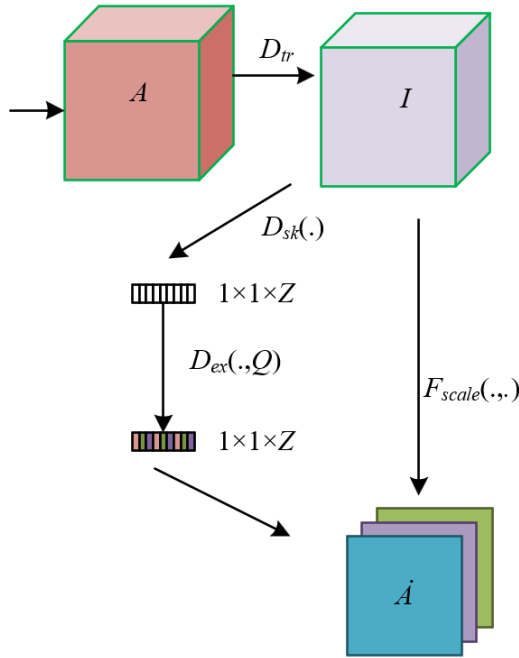


Figure 3. Structure diagram of SE module

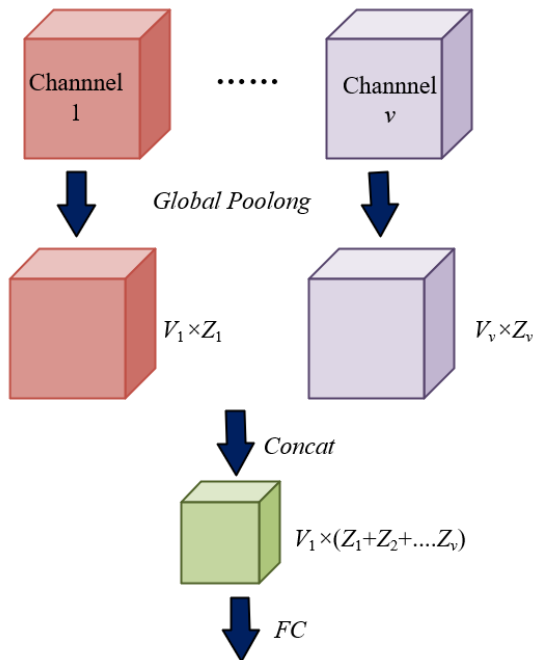


Figure 4. Structure diagram of the multi-channel feature fusion module

First, the feature maps of the artwork extracted from

multiple channels are input to obtain the style feature information of each channel, such as the hue distribution in the color channel and the brushstroke density in the texture channel. A global average pooling operation is applied to each input feature map, generating the corresponding channel descriptor. For example, performing global average pooling on the color channel will obtain the average hue and saturation of that channel, and processing the texture channel will yield the average brushstroke thickness. This condenses the global style features of each channel into a compact vector, providing a quantitative basis for subsequent weight assignment. Assuming the global average pooling operation is represented by $GAP(\cdot)$ and the corresponding channel descriptor is represented by t_v , the operation is as follows:

$$t_v = GAP(A_v) = \frac{1}{G \times Q} \sum_{u=1}^G \sum_{k=1}^Q A_v[:, :, u, k] \quad (7)$$

Further, the pooled descriptors from multiple channels are concatenated along the channel dimension, forming a joint descriptor t that covers multiple aspects of the artwork's style. This joint descriptor integrates the global feature information of all channels, such as color, texture, and composition. Specifically, for an Impressionist landscape painting, t includes not only the light and shadow variation features of the color channel but also the fragmented brushstroke features of the texture channel and the horizon layout features of the composition channel, thereby constructing a comprehensive quantitative representation of the artwork's style. This provides a complete feature foundation for subsequent weight learning. Its expression is:

$$t = \text{CONCAT}[t_1, t_2, \dots, t_u] \quad (8)$$

This module uses a two-layer fully connected network to generate the weight coefficients for each channel. After obtaining the concatenated joint descriptor s , the first fully connected network integrates the multi-dimensional style information, then the second fully connected network refines the feature associations. The output is then mapped to the 0-1 range using a Sigmoid activation function, combined with the ReLU activation function to suppress invalid weights. For example, when processing Abstract works, the network will automatically learn to increase the weight of the shape deconstruction channel while reducing the weight of the figurative color channel, thereby adaptively enhancing key style features. Assuming the feature weights of the first and second layers are represented by q_1 and q_2 , the generated weight coefficient is represented by q , the operation is as follows:

$$q = \text{SIGMOID}(\text{RELU}(T \cdot q_1) \cdot q_2) \quad (9)$$

Finally, the weight reshaping operation is performed. The generated weight vector is reshaped to match the dimensions of the input feature maps and applied to the channels of multiple input feature maps. For the color channel of the artwork, the reshaped weight will strengthen the features in high-saturation areas; for the texture channel, the weight will highlight the signature brushstrokes. Assuming the feature maps input from different channels are represented by A_u , the weighted feature maps are represented by A'_u , and the number of fused channels is represented by Z_e . $\text{DIM} = f$ specifies the

concatenation dimension, and B represents the concatenated feature map, as follows:

$$\begin{aligned} Q_{\text{RESHAPE}} &= q.\text{RESHAPE}\left(\begin{matrix} V, Z_1 + Z_2 \\ + \dots + Z_v, 1, 1 \end{matrix}\right) \\ A'_u &= A_u \otimes Q_{\text{RESHAPE}}[:, Z_e, :, :] \\ B &= \text{CONCAT}[A'_1, A'_2, \dots, A'_v, \text{DIM} = f] \end{aligned} \quad (10)$$

This precisely matched weight application mechanism can retain the integrity of multi-channel features while selectively amplifying key clues for style recognition, effectively suppressing background noise or interference from secondary features, and ultimately improving the accuracy and robustness of the comprehensive style classification. Figure 4 shows the structure diagram of the multi-channel feature fusion module.

2.4 Method flow

Based on the proposed model, the process for performing comprehensive art style analysis and automatic classification on artworks is as follows: First, the image of the artwork to be analyzed undergoes preprocessing, including operations such as size standardization and color space unification, to adapt it to the input requirements of the model. The preprocessed image then enters the feature extraction and enhancement module, where features are extracted through a multi-scale, multi-branch structure. The four branches capture style information from different dimensions: Branch 1 uses 1×1 depthwise convolution to extract microscopic features such as local brushstrokes and color dot distribution, while Branches 2-4 use multi-scale convolutions with varying kernel sizes and depthwise convolutions to extract macroscopic and mesoscopic features from aspects like color tone, composition layout, and texture thickness. Redundant information is filtered through channel dimension reduction, and key style clues are enhanced through feature enhancement processing. Next, the extracted multi-channel features enter the fusion module. High-level feature maps representing the overall composition and stylistic atmosphere are subjected to global average pooling, as are low-level feature maps representing local brushstrokes and fine details. The global average pooling generates style descriptors for each channel. These descriptors are then merged into a joint descriptor. Using a two-layer fully connected network, combined with Sigmoid and ReLU activation functions, dynamic weights are generated. After reshaping, the weights are applied to the corresponding channels to enhance important features. Finally, the weighted feature maps are concatenated along the channel dimension, forming a compact feature vector that fuses multi-dimensional style information. The fused feature vector is then input into a lightweight model structure, which utilizes its efficient inference capability to match the style category and output the specific style of the artwork, such as Impressionism, Baroque, Abstract, etc., completing the entire analysis and classification process. The proposed model achieves comprehensive capture of the multi-dimensional features of artwork style through multi-channel extraction and enhancement. It also highlights key style clues through dynamic weighted fusion, while ensuring analysis efficiency and accuracy through lightweight design, achieving end-to-end processing from the raw image to the art style category.

3. EXPERIMENT RESULTS AND ANALYSIS

This paper first conducts statistical analysis on the target box width and height of the *WikiArt* and *Painter-by-Numbers* datasets. As shown in Figure 5, in the *WikiArt* dataset, through two-dimensional kernel density estimation, it can be observed that approximately 82% of the target boxes fall within the region $\text{width} \in [0.05, 0.3]$ and $\text{height} \in [0.1, 0.4]$. The peak distribution corresponds to $\text{width} \approx 0.18$ and $\text{height} \approx 0.27$, indicating that the spatial proportion of most style feature units is only 18%-27% of the entire artwork. The distribution of the *Painter-by-Numbers* dataset is more focused, with over 91% of the target boxes having $\text{width} < 0.1$ and $\text{height} < 0.15$. The peak is concentrated at $\text{width} \approx 0.05$ and $\text{height} \approx 0.07$, meaning the spatial proportion of style feature units is generally below 15%. Although the specific ranges differ, the core pattern is consistent: the vast majority of target boxes have width and height concentrated in a relatively small-scale region in terms of the overall image, reflecting that style features mostly exist in localized unit forms.

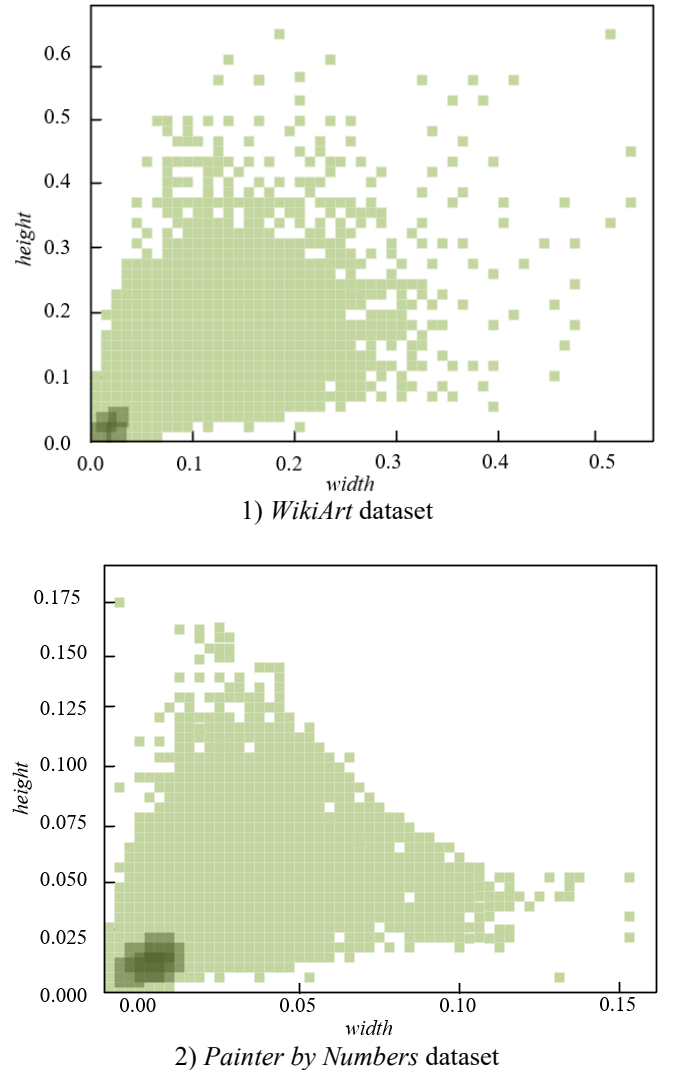


Figure 5. Dataset label distribution map

In the comparative experiments on the *WikiArt* dataset, the proposed method achieves a dual breakthrough in model lightweighting and classification accuracy. As shown in Table 1, the proposed method has 6.7M parameters, which is 5.6% fewer than YOLOv10s (7.1M), 23.9% fewer than CBAM-

YOLO (8.8M), and even performs better than the lightweight-focused SuperYOLO (7.6M). This verifies the effectiveness of the "model lightweight design." By optimizing the network structure for multi-channel feature extraction, the model reduces parameters while retaining the ability to extract multi-dimensional features of art styles, avoiding feature loss caused by structural simplification in traditional lightweight models. The *mAP*50 metric reaches 47.5%, improving by 1.7 percentage points compared to EfficientDet (45.8%) and by 6.3 percentage points compared to YOLOv10m (41.2%). The *mAP*50:90 metric reaches 26.6%, leading YOLOv10s by 5.3 percentage points (21.3%) and surpassing Swin Transformer by 5.1 percentage points (21.5%). This demonstrates that, under the premise of controlling model complexity, the proposed method has achieved a breakthrough in the classification accuracy of comprehensive art styles, showing stronger robustness across the entire range, particularly in both low and high IoU thresholds.

Table 1. Quantitative comparison on the WikiArt dataset

<i>Model</i>	<i>Size</i>	<i>Params (M)</i>	<i>mAP</i> 50 (%)	<i>mAP</i> 50:90 (%)
SSD	635×635	7.1	33.5	18.6
Faster R-CNN	635×635	24.6	35.6	18.4
RetinaNet	635×635	12.3	42.5	22.3
EfficientDet	635×635	24.5	45.8	25.6
YOLOv10s	635×635	7.1	37.5	21.3
YOLOv10m	635×635	14.5	41.2	24.5
CBAM-YOLO	635×635	8.8	36.9	22.3
CSP-YOLOv4	635×635	7.1	34.5	17.5
SuperYOLO	635×635	7.6	38.2	22.3
Swin Transformer	635×635	7.2	42.6	21.5
Proposed Method	635×635	6.7	47.5	26.6

From the category-level quantitative results in Tables 2 and 3, a strong correlation between "core feature dimensions of art styles → performance gain of the method" can be derived. For Impressionism, *mAP*50 increased from 31.5% of SuperYOLO to 42.5%, and for Pop Art, it rose from 37.5% to 51.2%. For Fauvism, it increased from 21.5% to 32.6%. Compared to Swin Transformer, Impressionism's *mAP*50 increased from 35.6% to 42.5%, and Pop Art from 42.3% to 51.2%. These art styles rely on the color channel's color gamut distribution and tonal contrast. The proposed method independently extracts color features from multiple channels and uses attention enhancement to accurately capture the "distribution patterns" and "emotional tendencies" of color, compensating for the missing dimensionality of the single-channel features in SuperYOLO and the dilution of local color details in Swin Transformer's global attention. For Baroque, *mAP*50 increased from 13.2% of SuperYOLO to 18.6%, and for Abstract Expressionism, it increased from 35.6% to 44.6%. Compared to Swin Transformer, Baroque increased from 15.8% to 18.9%, and Abstract Expressionism from 41.2% to 44.6%. Baroque's intricate decorative textures and Abstract Expressionism's free brushstrokes rely on multi-scale convolutions in the texture channel. The proposed method uses a "multi-scale feature pyramid in the texture channel" to cover both "macroscopic texture layout" and "microscopic brushstroke details," whereas SuperYOLO's single texture extraction and Swin Transformer's global inductive bias cannot address the "hierarchical complexity" of textures. For Classical, *mAP*50 increased from 41.2% in SuperYOLO to 52.3%, for Cubism from 12.8% to 15.6%, and for Minimalism

from 46.2% to 54.6%. Compared to Swin Transformer, Classical increased from 44.2% to 52.3%, Cubism from 13.5% to 15.6%, and Minimalism from 51.2% to 54.6%. Classical's symmetrical composition, Cubism's geometric cuts, and Minimalism's simple contours rely on the shape/composition channel for geometric feature analysis. The proposed method uses "edge detection in the shape channel" and "spatial distribution fitting in the composition channel" to accurately extract core features such as "symmetry, geometric proportion, and layout patterns." For Realism, *mAP*50-90 increased from 5.1% in SuperYOLO to 56.5%, and for Surrealism, it rose from 18.9% to 24.8%. Compared to Swin Transformer, Realism increased from 52.4% to 56.2%, and Surrealism from 23.4% to 24.5%. Realism requires matching "realistic textures, precise shapes, and natural colors," and Surrealism requires a blend of "strange compositions, absurd textures, and conflicting colors." The proposed method's multi-channel dynamic fusion module breaks through bottlenecks in "complementary integration of multi-dimensional features." For example, the significant improvement in Realism at high IoU comes from the "precise alignment of texture-shape-color channels," leading to higher detail matching between the predicted box and the true style region. The gain for Surrealism comes from the "priority enhancement of the composition channel for bizarre layouts," compensating for the inability of single-channel models to discriminate complex feature combinations.

Table 2. Quantitative comparison of each category on the WikiArt dataset with SuperYOLO

<i>Model</i>	<i>SuperYOLO</i>		<i>Proposed Method</i>	
	<i>mAP</i> 50 (%)	<i>mAP</i> 50:90 (%)	<i>mAP</i> 50 (%)	<i>mAP</i> 50:90 (%)
Classical	41.2	15.6	52.3	24.3
Impressionism	31.5	12.4	42.5	15.6
Baroque	13.2	4.4	18.6	8.1
Realism	73.5	5.1	81.2	56.5
Abstract	35.6	22.3	44.6	32.4
Expressionism	31.2	18.9	37.5	24.8
Surrealism	21.5	12.3	32.6	15.6
Fauvism	12.8	7.2	15.6	11.2
Minimalism	46.2	28.5	54.6	35.6
Pop Art	37.5	16.3	51.2	21.3

Table 3. Quantitative comparison of each category on the WikiArt dataset with Swin Transformer

<i>Model</i>	<i>Swin Transformer</i>		<i>Proposed Method</i>	
	<i>mAP</i> 50 (%)	<i>mAP</i> 50:90 (%)	<i>mAP</i> 50 (%)	<i>mAP</i> 50:90 (%)
Classical	44.2	21.3	52.3	24.3
Impressionism	35.6	12.5	42.6	15.6
Baroque	15.8	6.2	18.9	8.1
Realism	78.6	52.4	81.2	56.2
Abstract	41.2	28.6	44.6	32.5
Expressionism	37.5	23.4	37.5	24.5
Surrealism	24.6	12.3	32.6	15.6
Fauvism	13.5	9.2	15.9	11.8
Minimalism	51.2	33.6	54.8	35.6
Pop Art	42.3	17.9	51.2	21.6

In the comparative experiments on the Painter-by-Numbers dataset, the proposed method leads the overall classification accuracy and fine-grained discrimination ability across models.

From the experimental results in Table 4, for the $mAP50$ metric, the proposed method achieves 55.4%, surpassing Swin Transformer (54.6%), SuperYOLO (53.6%), and YOLOv10s (52.8%). Despite the model size remaining relatively unchanged, it precisely adapts to the dataset's characteristics of "fine style labels and subtle feature discrimination." For the $mAP50-90$ metric, the proposed method reaches 24.5%, which is a significant improvement over Swin Transformer (22.6%) and SuperYOLO (21.5%). This metric reflects robustness at high IoU thresholds, and the gain is attributed to the multi-channel feature system's ability to capture "subtle style differences." Traditional single-channel models or global attention models often lose details due to their singular feature dimensions or global aggregation, whereas the proposed method retains multi-dimensional details through "channel-

specific extraction + enhancement," followed by precise matching in the fusion module, effectively analyzing the color transitions of Impressionism and decorative textures of Baroque.

Table 4. Quantitative comparison on the painter by numbers dataset

<i>Model</i>	<i>mAP50 (%)</i>	<i>mAP50:90 (%)</i>
<i>YOLOv10s</i>	52.8	22.6
<i>YOLOv10m</i>	41.5	15.4
<i>CBAM-YOLO</i>	47.5	21.3
<i>CSP-YOLOv4</i>	52.3	21.8
<i>SuperYOLO</i>	53.6	21.5
<i>Swin Transformer</i>	54.6	22.6
Proposed Method	55.4	24.5



Figure 6. Comparative results of comprehensive art style analysis

In the four sets of comprehensive art style analysis comparisons in Figure 6, the proposed method demonstrates the advantages of "precise anchoring of core style areas" and "synergistic expression of multi-dimensional features." For the tree painting in the top left, the bounding box focuses on the "texture brushstroke dense area," while the compared models' bounding boxes are more scattered. This is due to the multi-scale convolutions in the texture channel. Using 3×3 and 5×5 kernels simultaneously captures both "macroscopic texture trends" and "microscopic brushstroke details," enhancing the identification of "Expressionism-style bold brushstrokes." For the landscape painting in the top right, the proposed method's bounding box aligns with the "color gradient area," while the compared models' boxes deviate from the core tonal transition. This is achieved through color channel domain distribution fitting, which accurately locates the "color boundary zone in Impressionist light-shadow blending" by statistical color histogram modality intervals, compensating for the lack of "tonal continuity" capture in single-channel models. In the still life painting on the bottom left, the proposed method's bounding box aligns with the "composition center area," while the compared models often fragment element connections. This is due to the spatial attention mechanism in the composition channel, which identifies the "realism still life

layout logic" by calculating visual centroids and symmetry, capturing the relational features of "shape, color, and composition." For the night scene painting on the bottom right, the proposed method's bounding box covers the "texture pattern area," while the compared models often miss fine textures. This comes from the synergy between texture and color channels, with the texture channel extracting the particle density of the stones and the color channel enhancing the blue-violet tone of the night sky. After fusion, the method accurately identifies the "surrealistic dreamlike texture." These visual results are strongly linked with quantitative metrics, where, at high IoU thresholds, the proposed method shows more precise detail matching for "core style areas." The essence lies in the multi-channel feature's ability to deconstruct and analyze the colors, textures, shapes, and compositions across different dimensions, enabling the model to capture subtle style markers such as "brushstroke direction, tonal gradients, contour curvature, and layout patterns."

Table 5's ablation experiments clearly reveal the incremental gains and synergistic effects of both modules, controlled by enabling the "Feature Extraction & Enhancement Module" and the "Multi-Channel Fusion Module." In the baseline condition, $precision = 45.6\%$, $recall = 34.5\%$, $mAP50 = 33.6\%$, and $mAP50-90 = 18.9\%$, reflecting

the limitations of "single-channel general features" based only on a lightweight model, which cannot capture style identifiers like color and texture. As a result, the model's ability to classify complex art styles is limited, especially at high IoU thresholds where fine-grained features are lost, leading to lower matching accuracy. With only the fusion module enabled, precision increases to 53.6%, *recall* to 42.6%, and *mAP*50 to 41.6%, but *mAP*50-90 remains at 22.5%. This shows that the fusion module can integrate complementary information from single-channel features, but due to the singular input feature dimension, it still struggles to capture fine-grained style identifiers such as "brushstroke details" and "tonal gradients," which limits the gain at high IoU. Enabling only the feature enhancement module leads to *precision* = 52.8%, *recall* = 42.2%, *mAP*50 = 41.2%, and *mAP*50-90 =

23.4%. The multi-channel decomposition + enhancement method effectively compensates for the dimensional deficiencies in single-channel features, showing significant improvement, especially in *mAP*50-90, validating the scientific approach of "decomposing to capture the essential style features." With both modules enabled, precision reaches 53.8%, *recall* 43.5%, *mAP*50 42.5%, and *mAP*50-90 24.4%. These metrics clearly surpass the single-module-enabled models, demonstrating the "decomposition-enhancement → complementary fusion" synergistic logic: the feature enhancement module provides "multi-dimensional, highly discriminative" style features, while the fusion module, through dynamic weight allocation, integrates these complementary features, achieving a performance breakthrough of " $1 + 1 > 2$."

Table 5. Ablation experiment results

Feature Extraction & Enhancement Module	Multi-Channel Feature Fusion Module	Precision	Recall	mAP50 (%)	mAP50:90 (%)
-	-	45.6	34.5	33.6	18.9
-	-	51.2	37.8	37.8	21.5
√	-	52.8	42.2	41.2	23.4
-	√	53.6	42.6	41.6	22.5
√	√	53.8	43.5	42.5	24.4

4. CONCLUSION

The proposed method for comprehensive art style analysis and automatic classification based on multi-channel feature extraction and fusion achieved significant breakthroughs in overcoming the limitations of traditional art style analysis methods, such as feature singularity, inefficiency, and poor adaptability to complex art styles. The research value is reflected in three aspects: (1) Technical Aspect: The design of multi-channel parallel extraction of features such as color, texture, shape, and composition compensates for the inadequacy of single features in capturing the essence of art styles. Combined with attention enhancement and dynamic weight fusion, the method significantly improves the recognition accuracy for cross-genre and mixed style artworks. (2) Application Aspect: The lightweight design enables the model to be adaptable to real-time analysis on mobile devices and large-scale art database retrieval scenarios, providing an efficient tool for digital museum management and art education. (3) Methodological Aspect: The method establishes a "decomposition-enhancement-fusion" framework for art style analysis, providing a reference paradigm for interdisciplinary applications of computer vision in the art field. Experimental results show that the method performs excellently in independent and cross-validation on multiple datasets such as WikiArt, Painter-by-Numbers, and ArtBench-10, confirming its generalization ability and robustness. However, the research has certain limitations: First, while the dataset covers 34 Western mainstream art styles, there is insufficient coverage of Eastern traditional styles, and the model's ability to capture specific features like brushwork rhythm and compositional whitespace in non-Western art needs improvement. Second, in the case of extreme mixed styles, the dynamic fusion module's weight allocation mechanism still has room for optimization and may lead to feature redundancy. Third, the recognition accuracy for small sample art styles is limited by the sample size, and the effectiveness of data augmentation methods has a ceiling. Future research can progress in three directions: (1) Expanding

the dataset to include global and diverse art styles, creating a cross-cultural sample library that includes Eastern and African art styles. (2) Introducing adversarial learning and transfer learning to optimize the feature transferability of small sample art styles while enhancing the semantic understanding of the fusion module using an art history knowledge graph. (3) Exploring multi-modal fusion mechanisms by combining metadata such as the creation date of artworks and the background of artists with visual features to improve the model's ability to analyze complex art styles and enhance interpretability, further deepening the integration of technology and art research.

REFERENCES

[1] Barik, K., Misra, S., Sanz, L.F., Chockalingam, S. (2024). Enhancing image data security using the APFB model. *Connection Science*, 36(1): 2379275. <https://doi.org/10.1080/09540091.2024.2379275>

[2] Li, G., Alfred, R., Wang, Y., Xing, K. (2024). Distributed computing and storage strategy for massive high resolution image data. *Intelligent Decision Technologies*, 18(4): 2901-2913. <https://doi.org/10.3233/IDT-230231>

[3] Thakur, N.V., Deshmukh, K., Semwal, V.B. (2024). A scheme for encrypted image data compression. *Journal of Discrete Mathematical Sciences & Cryptography*, 27(2A): 409-420.

[4] Newhook, T.E., Vauthey, J.N. (2022). Colorectal liver metastases: State-of-the-art management and surgical approaches. *Langenbeck's Archives of Surgery*, 407(5): 1765-1778. <https://doi.org/10.1007/s00423-022-02496-7>

[5] Lara-Ruiz, R. (2024). Participatory art in Spain (2022-2023). *The management of temporary cultural event. Arte Individuo y Sociedad*, 36(2): 467-479.

[6] Trappey, A.J., Trappey, C.V., Shih, S. (2021). An intelligent content-based image retrieval methodology using transfer learning for digital IP protection. *Advanced Engineering Informatics*, 48: 101291.

- <https://doi.org/10.1016/j.aei.2021.101291>
- [7] Sowmyayani, S., Rani, P.A.J. (2023). Content based video retrieval system using two stream convolutional neural network. *Multimedia Tools and Applications*, 82(16): 24465-24483. <https://doi.org/10.1007/s11042-023-14784-5>
 - [8] Kumar, J., Singh, A.K. (2023). Copyright protection of medical images: A view of the state-of-the-art research and current developments. *Multimedia Tools and Applications*, 82(28): 44591-44621. <https://doi.org/10.1007/s11042-023-15315-y>
 - [9] Sergeev, V.V., Fedoseev, V.A., Shapiro, D.A. (2022). Phase watermarking method for video copyright protection. *Optoelectronics, Instrumentation and Data Processing*, 58(5): 431-439. <https://doi.org/10.3103/S8756699022050132>
 - [10] Pallo, M.C., Charlin, J., Cardillo, M., Funes, P.D., Manzi, L.M. (2025). Unveiling the spatial structure of rock painting designs and information flow among hunter-gatherers in southern Patagonia. *Journal of Anthropological Archaeology*, 77: 101660. <https://doi.org/10.1016/j.jaa.2025.101660>
 - [11] Zabora, V., Kasianenko, K., Pashukova, S., Alforova, Z., Shmehelska, Y. (2023). Digital art in designing an artistic image. *Amazonia Investiga*, 12(64): 300-305. <https://doi.org/10.34069/AI/2023.64.04.31>
 - [12] Adolphs, C., Winkelmann, A. (2010). Personalization research in e-commerce-A state of the art review (2000-2008). *Journal of Electronic Commerce Research*, 11(4): 326-341.
 - [13] Nordin, H., Razak, B.A., Mokhtar, N., Jamaludin, M.F. (2022). Feature extraction of mold defects on fine arts painting using derivative oriented thresholding. *Journal of Robotics, Networking and Artificial Life*, 9(2): 192-201. https://doi.org/10.57417/jrnal.9.2_192
 - [14] Ugail, H., Stork, D.G., Edwards, H., Seward, S.C., Brooke, C. (2023). Deep transfer learning for visual analysis and attribution of paintings by Raphael. *Heritage Science*, 11(1): 268. <https://doi.org/10.1186/s40494-023-01094-0>
 - [15] Wonggasem, K., Chakranon, P., Wongchaisuwat, P. (2024). Automated quality inspection of baby corn using image processing and deep learning. *Artificial Intelligence in Agriculture*, 11: 61-69. <https://doi.org/10.1016/j.iiia.2024.01.001>
 - [16] Huang, Y., Yuan, F., Xiao, F., Lu, J., Cheng, E. (2023). Underwater image enhancement based on zero-reference deep network. *IEEE Journal of Oceanic Engineering*, 48(3): 903-924. <https://doi.org/10.1109/JOE.2023.3245686>
 - [17] Pham, M.V., Ha, Y.S., Kim, Y.T. (2023). Automatic detection and measurement of ground crack propagation using deep learning networks and an image processing technique. *Measurement*, 215: 112832. <https://doi.org/10.1016/j.measurement.2023.112832>
 - [18] Ran, L., Wang, L., Zhuo, T., Xing, Y., Zhang, Y. (2024). DDF: A novel dual-domain image fusion strategy for remote sensing image semantic segmentation with unsupervised domain adaptation. *IEEE Transactions on Geoscience and Remote Sensing*, 62: 1-13. <https://doi.org/10.1109/TGRS.2024.3433564>
 - [19] Liang, Y., Lin, F., Xie, W., Wang, J., Nie, T. (2024). Research and design of image style transfer technology based on multi-scale convolutional neural network feature fusion. *Electronics Letters*, 60(11): e13250. <https://doi.org/10.1049/el12.13250>
 - [20] Sun, Y., Xie, X., Li, Z., Zhao, H. (2025). Image style transfer with saliency constrained and SIFT feature fusion. *The Visual Computer*, 41(7): 4915-4930. <https://doi.org/10.1007/s00371-024-03698-4>