

Article

An Explainable AI Classification Framework for Impressionist Paintings Based on Fourier Transform Feature Disentanglement: Toward Human–AI Collaborative Art Appreciation

Cunyuan Yang ^{1,2} , Ken Nah ^{2,3} , Zhe Qian ^{4,*}  and Guangliang Sang ^{2,3,*} ¹ School of Creative Arts and Design, Zhejiang Polytechnic University of Mechanical and Electrical Engineering, Hangzhou 310053, China; yangcunyuanyuan@zime.edu.cn² International Design Trend Center, Hongik University, Seoul 04068, Republic of Korea; knah@hongik.ac.kr³ College of Arts and Physical Education, Kangnam University, Yongin 16979, Republic of Korea⁴ Zhejiang Gongshang University Hangzhou College of Commerce, Hangzhou 311508, China

* Correspondence: 2030027@zjhzc.edu.cn (Z.Q.); glsang@mail.hongik.ac.kr (G.S.)

Abstract

Impressionist artworks, marked by distinctive brushstrokes and light effects, pose challenges for holistic deep learning models. Existing classification methods often lack interpretability and mathematical rigor. To address this, we propose an explainable AI framework for Impressionist painting classification based on feature disentanglement. The framework uses Fourier transform to decompose images into magnitude and phase spectra and then applies frequency filtering to separate high-frequency components (brushstroke textures, edges) from mid-to-low frequencies (global composition, color transitions). A state space model captures long-range dependencies within frequency representations. Disentangled features are fused with the original image features via a dual-branch injection mechanism. Experiments show that a lightweight model with 10M parameters, using this injection strategy, outperforms conventional baselines with 25M parameters. Moreover, the approach provides interpretability by explicitly linking classification decisions to brushstroke organization, frequency energy distribution, and semantic composition. This work offers a unified, computationally efficient, and mathematically rigorous solution for feature disentanglement in artistic image analysis, laying a theoretical foundation for explainable computational aesthetics and human–AI collaborative appreciation.

Keywords: impressionist painting classification; feature disentanglement; frequency-domain analysis; explainable AI; state space models; computational aesthetics

MSC: 42A38



Academic Editor: Junming Chen

Received: 24 June 2026

Revised: 17 August 2026

Accepted: 18 August 2026

Published: 29 August 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

1.1. Challenges in the Classification of Impressionist Paintings

Emerging in France during the 1870s, Impressionism fundamentally transformed Western painting by separating the depiction of light and color from traditional representations of contour and volume. Artists such as Claude Monet, Pierre-Auguste Renoir, Vincent van Gogh, and Paul Gauguin reconstructed visual expression through explicit brushstrokes, complementary color juxtaposition, and the depiction of transient lighting conditions [1,2]. This deliberate decomposition of visual elements constitutes one of the defining characteristics of Impressionist art.

We deliberately focus on Impressionism because its systematic internal technical consistency—the shared grammar of separated brushwork and optical mixing—provides a necessary controlled environment for validating frequency-based feature disentanglement. An in-depth diagnostic evaluation within a homogeneous stylistic framework offers greater methodological rigor than achieving high accuracy on a heterogeneous dataset.

However, such characteristics present a fundamental challenge to contemporary deep learning-based visual recognition systems. Conventional convolutional neural networks (CNNs) and Vision Transformers (ViTs) generally encode images as unified semantic representations by jointly modeling texture, color, and structural information through end-to-end learning. The resulting high-dimensional latent embeddings implicitly fuse heterogeneous visual cues into holistic representations.

For artistic styles whose visual identity is intrinsically constructed through the separation of multiple perceptual dimensions, such as Impressionism, holistic feature learning may not only be suboptimal but may also obscure the essential stylistic mechanisms that distinguish the movement itself. In other words, the internal visual logic of Impressionist paintings is inherently disentangled, whereas current recognition models typically assume entangled visual representations.

Recent studies on AI-based artwork classification have achieved considerable progress through supervised learning using CNN and Transformer architectures [3,4]. Although classification accuracy has steadily improved, these approaches provide limited interpretability regarding why a painting is assigned to a particular artistic category. Existing models simply learn correlations between labeled images and style classes, without identifying the brushstroke configurations, illumination patterns, color arrangements, or compositional strategies that contribute to the final decision. Such a lack of interpretability severely limits their applicability in downstream scenarios including art education, museum interpretation, and digital cultural heritage preservation.

1.2. Feature Disentanglement as a Paradigm for Explainable Artificial Intelligence

Feature disentanglement has emerged as an important paradigm in computer vision and generative modeling for separating heterogeneous visual factors into independently interpretable representations. Among existing approaches, Fourier transform provides a mathematically rigorous and computationally efficient framework for decomposing paintings into frequency-domain components.

By transforming an image from the spatial domain into the frequency domain, Fourier analysis separates visual information according to different frequency distributions. Brushstroke textures, pigment accumulation, and canvas structures exhibit distinct spectral characteristics, enabling their quantitative separation through frequency filtering techniques. High-pass filters can suppress slowly varying low-frequency components while enhancing high-frequency brushstroke details, whereas notch filtering can isolate periodic canvas textures and emphasize artist-specific stroke patterns.

Previous studies have demonstrated the effectiveness of Fourier-domain analysis for artistic image understanding [5]. Enhanced brushstroke visibility while suppressing canvas structures through Fourier filtering, combined Fast Fourier Transform with convolutional neural networks to analyze different painting techniques. These studies suggest that frequency-domain decomposition provides a mathematically interpretable representation for quantitative brushstroke analysis and artistic style understanding.

Viewing feature disentanglement as a design principle for Impressionist painting classification offers several potential advantages. First, it transforms abstract concepts such as artistic style and visual atmosphere into observable and quantifiable computational representations. Second, unlike conventional approaches that focus primarily on enabling

machines to learn more features automatically, the proposed paradigm emphasizes enabling machines to perceive paintings through a process more closely aligned with human visual appreciation. Third, from a human–computer interaction perspective, disentangled representations establish interpretable and transferable channels that support interactive explanation and collaborative art appreciation.

1.3. Research Questions and Contributions

Based on the above observations, this study addresses the following research questions:

RQ1: How can the visual characteristics of Impressionist paintings be systematically decomposed into physical, stylistic, and semantic feature layers?

RQ2: How can existing image decomposition and latent-space disentanglement techniques be integrated into an explainable AI framework for Impressionist painting classification?

RQ3: What design implications does such a framework provide for human–AI collaborative art appreciation and interactive visual analysis?

To answer these questions, this study makes three primary contributions.

First, it proposes a three-layer feature disentanglement framework that integrates heterogeneous computational techniques into a unified architecture for artistic image classification.

Second, it presents a complete research protocol covering methodological design, dataset construction, feature extraction strategies, and evaluation metrics for interpretable artwork classification.

Third, from the perspectives of design research and human–computer interaction, it discusses how disentangled visual representations can support interpretable interaction, collaborative artistic understanding, and future intelligent art appreciation systems.

2. Related Work

2.1. AI-Based Artwork Classification

Over the past decade, research on AI-assisted artwork classification has evolved along three major directions. The first focuses on art movement and style classification, where large-scale annotated datasets such as WikiArt are employed to train deep neural networks for predicting artistic styles, the authorship, and historical periods [3,4]. Representative approaches typically adopt convolutional neural networks (CNNs) or Vision Transformers (ViTs) as backbone architectures and formulate artwork understanding as a supervised image classification task.

The second research direction concerns art retrieval and recommendation systems, which aim to construct semantically meaningful embedding spaces capable of supporting cross-modal retrieval and artwork recommendation [6,7]. Rather than emphasizing classification accuracy alone, these methods focus on semantic similarity learning and multimodal representation alignment between visual and textual information.

The third direction involves AI-assisted artistic creation, where generative models such as Generative Adversarial Networks (GANs) and diffusion models have demonstrated remarkable capability in artistic style synthesis and image generation [7,8]. These approaches primarily investigate controllable image generation and style transfer through latent representation learning.

Despite their methodological differences, these three research streams share a common characteristic: they predominantly rely on end-to-end deep neural architectures that implicitly encode artistic knowledge within high-dimensional latent spaces. Although explainability has recently attracted increasing attention in Explainable Artificial Intelligence (XAI), relatively few studies have treated feature disentanglement as a fundamental design principle for artwork classification. Consequently, existing systems remain limited in their ability to explain which specific artistic characteristics contribute to their classification decisions.

2.2. Frequency-Domain Feature Disentanglement and Brushstroke Representation

Recent advances in computational aesthetics and style transfer have increasingly explored frequency-domain representations as a mathematically interpretable approach to artistic style analysis. Among these methods, the Fourier Transform provides a natural mechanism for decomposing visual information into different frequency components, thereby enabling the separation of global appearance and local structural characteristics.

Specifically, the amplitude spectrum primarily describes global visual properties such as color distribution, illumination variation, and contrast, whereas the phase spectrum preserves structural information including object boundaries, geometric arrangements, and edge configurations. This amplitude–phase decomposition establishes a theoretical foundation for independently modeling artistic style and image content [9].

Building upon this principle, researchers systematically investigated multiple feature decomposition techniques—including Fast Fourier Transform (FFT), Discrete Cosine Transform (DCT), Principal Component Analysis (PCA), and Independent Component Analysis (ICA)—to represent and manipulate artistic styles computationally. Their framework demonstrated that stylistic elements can be separated, recombined, and manipulated through mathematically interpretable transformations.

Similarly, studies analyzed Fourier power spectra across various artistic images and observed that paintings exhibit scale-invariant $1/f^2$ spectral statistics, together with stronger isotropic characteristics than natural photographs. These findings suggest that different artistic styles possess distinguishable frequency-domain signatures that can serve as effective computational descriptors.

Beyond style representation, transforming frequency-domain information into interpretable brushstroke semantics has become an important challenge for controllable artistic generation. Existing studies have attempted to model stroke trajectories, orientation distributions, and texture organizations through handcrafted spatial representations or brushstroke rendering techniques. Nevertheless, most existing approaches remain confined to spatial-domain analysis.

A critical unresolved problem is the systematic mapping of spectral properties—directional energy, spectral density, and frequency anisotropy—onto human-interpretable artistic concepts, including brushstroke granularity, orientation, density, and organizational structure. Addressing this issue would establish a direct connection between Fourier analysis and explainable artistic representation learning.

Overall, frequency-domain decomposition provides a mathematically rigorous foundation for computational modeling of artistic styles. Its inherent ability to separate structural and stylistic information offers promising opportunities for developing more interpretable and controllable AI systems for artwork analysis and generation.

2.3. Research Gap: From Engineering Optimization to Design-Oriented Deconstruction

Most existing deep learning approaches to painting classification primarily pursue improvements in recognition accuracy through deeper architectures, attention mechanisms, or larger datasets [10]. Implicitly, these engineering-oriented methods assume that the optimal strategy is to allow neural networks to learn discriminative representations from entire images as black-box systems.

From a design perspective, however, artistic styles are not holistic entities but are instead composed of multiple design elements, including brushstroke morphology, chromatic organization, compositional rhythm, texture distribution, spatial balance, and visual hierarchy. These design elements jointly determine the perceptual identity of an artwork and should, therefore, constitute the fundamental units of computational analysis.

Interestingly, many of these visual components possess corresponding representations in the frequency domain. Low-frequency amplitudes encode global illumination and tonal organization [11], while high-frequency phase information preserves edges, textures, and brushstroke details [12]. Furthermore, power spectral density distributions across different frequency bands implicitly characterize stroke orientation, spatial scale, and texture complexity.

Consequently, a design-oriented classification framework should not simply learn unified latent representations but should instead explicitly decompose artistic images into multiple interpretable feature spaces, independently analyze different visual components, and integrate evidence across multiple perceptual dimensions to achieve more robust and explainable classification.

Despite the potential advantages of this perspective, relatively few studies have systematically incorporated design-oriented decomposition into AI-based painting classification. Existing frequency-domain approaches are typically employed for data augmentation or style transfer, while Fourier phase information is often used merely to improve recognition performance [13]. These methods remain fundamentally accuracy-driven and rarely address the more fundamental question of what design knowledge an AI system should perceive when understanding artworks.

Accordingly, three important research gaps can be identified.

First, the design-oriented definition of disentangled features remains underdeveloped. Existing studies have yet to establish a systematic taxonomy of computationally observable design elements—such as brushstroke granularity, directional entropy, hue partitioning, or frequency-domain anisotropy—and clarify their causal relationship with artistic style recognition.

Second, the mapping between frequency-domain features and design semantics remains largely unexplored. Although amplitude and phase spectra provide coarse-grained decomposition of style and structure, there is currently no mature framework capable of translating spectral characteristics into interpretable artistic modules such as coarse brushstroke regions, smooth transition areas, or texture organizations.

Third, explicit evidence fusion mechanisms for disentangled design representations remain absent. If multiple design dimensions are independently extracted, corresponding decision-level integration strategies should be designed to combine heterogeneous evidence rather than simply concatenating features within a unified latent embedding.

In summary, while existing engineering-oriented models have achieved remarkable improvements in classification accuracy, they often sacrifice interpretability by overlooking the internal design logic embedded within artistic images. Reconsidering what an AI system should perceive and how these perceptual elements should contribute to classification from a design-oriented perspective may open a new research direction that simultaneously enhances recognition performance and explainability. In this context, Fourier-based frequency-domain decomposition provides a natural computational foundation for achieving interpretable feature disentanglement in artistic image analysis.

3. Methodology

3.1. Dataset Construction

The dataset was constructed by sourcing artworks from publicly accessible online repositories. A total of 2000 paintings created by four representative Impressionist artists were initially acquired through automated web crawling. Duplicate images and low-quality samples were subsequently removed through manual inspection, and all remaining images were retained for further analysis.

Each artwork was independently labeled with its corresponding artist tag by a domain expert to ensure consistency and semantic correctness of the annotation. The final dataset

was randomly divided into a training set, a validation set, and a test set in a ratio of 80:10:10, resulting in 1600 training images, 200 validation images, and 200 test images. To demonstrate the effectiveness of style decoupling, we used the Gram style loss metric and LPIPS perceptual similarity, which quantify the feature relevance divergence between style and content, as shown in Table 1.

Table 1. Dataset composition details.

Style	Train Number	Test Number	Validation Number	Gram [10] ↓	LPIPS [14] ↓
Monet	440	55	55	0.000465	0.50349
Leano	440	55	55	0.000600	0.52107
Van Gogh	360	45	45	0.001289	0.45801
Gauguin	360	45	45	0.001472	0.47892
Total	1600	200	200	0.013411	0.49037

Copyright and Ethical Statement: All images included in the dataset correspond to artworks created by deceased artists whose copyrights have expired and entered the public domain. Consequently, no copyright disputes exist regarding the use of these materials for academic research purposes.

The data collection procedure complied with the robots protocols and terms of service of the target websites, and the crawling process did not interfere with their normal operation. The dataset contains no personal information or identifiable human subjects and is used exclusively for non-commercial scientific research.

3.2. Physical–Style Feature Disentanglement Module

The core component of the proposed framework is a Physical–Style Feature Disentanglement (PSFD) module, which aims to explicitly separate physical brushstroke characteristics from high-level semantic information through frequency-domain decomposition.

Unlike conventional end-to-end visual models that directly learn latent representations from raw images, the proposed module first transforms paintings into the frequency domain using the Fourier Transform, enabling mathematically interpretable feature decomposition. Our design primarily incorporates human involvement in the deep learning model, much like a teacher instructing a student. During model training, crucial information can be pre-injected into the model’s feature learning through Fourier transform filtering. This approach differs significantly from traditional deep learning and traditional Fourier transform methods, as shown in Table 2.

Table 2. The differences between the method presented in this paper and traditional Fourier transform and traditional deep learning.

Method Category	Traditional Fourier Style Analysis	End-to-End Depth Model	Ours Method
Core technologies	Hand-designed frequency domain filters and power spectrum statistics [15]	Black-box feature learning based on neural networks [16]	Feature style injection, independent feature learning module
Style Analysis	Based on global/local spectral statistics (such as radial spectrum, angular spectrum)	Global pooling or style loss for hidden activation	Dual-path learning enhances designer participation in model learning by incorporating features deemed necessary by the designer as branch learning components, thus increasing their weight in the end-to-end model.

Table 2. Cont.

Method Category	Traditional Fourier Style Analysis	End-to-End Depth Model	Ours Method
Main shortcomings	Fixed features make it difficult to capture complex, high-order stylistic semantics; reliance on hand-crafted features also limits adaptability to diverse artistic expressions.	Lacking inherent interpretability, relying on pre-trained weights, and unable to establish a connection with frequency domain art theory.	Designers need to have a certain basic understanding of aesthetics, such as understanding the main stylistic characteristics of a painting style as discussed in this article.
Representative work	[17,18]	[19–21]	[22,23]

First, perform the Two-Dimensional Discrete Fourier Transform (DFT) on the image:

$$F(u, v) = \sum_{x=0}^{H-1} \sum_{y=0}^{W-1} f(x, y) e^{-j2\pi(ux/W + vy/H)}$$

Here $f(x, y)$ represents the grayscale image size, and H, W are the image dimensions.

The resulting frequency spectrum contains both global structural information and localized texture distributions. Low-frequency components primarily encode illumination variation, color composition, and large-scale structural organization, whereas high-frequency components correspond to brushstroke boundaries, pigment accumulation, and fine texture details that are highly discriminative for Impressionist style recognition. To enhance these stylistic characteristics, a high-pass filtering operation is applied in the frequency domain. Execute the following operation with this.

Spectrum Centralization:

$$F_{shift}(u, v) = F(u - \frac{H}{2} \bmod H, v - \frac{W}{2} \bmod W)$$

Construct high-pass mask:

$$P(u, v) = \begin{cases} 0, & \text{if } \left|u - \frac{H}{2}\right| < R \text{ and } \left|v - \frac{W}{2}\right| < R \\ 1, & \text{otherwise} \end{cases}$$

Here, R represents the radius, and in this article, the parameter is set to 10.

Frequency domain filtering:

$$G_{shift}(u, v) = F_{shift}(u, v) \cdot P(u, v)$$

Anti-centralization:

$$G(u, v) = G_{shift}\left(u + \frac{H}{2} \bmod H, v + \frac{W}{2} \bmod W\right)$$

Invert the two-dimensional Discrete Fourier Transform (DFT), and take the actual part:

$$g(u, v) = \text{Re}(\mathcal{F}^{-1}G(u, v))$$

This operation suppresses redundant low-frequency information while preserving fine-grained artistic textures that contribute significantly to style discrimination, Reference for Fourier Transform High-Pass Filtering Effect Figure 1.

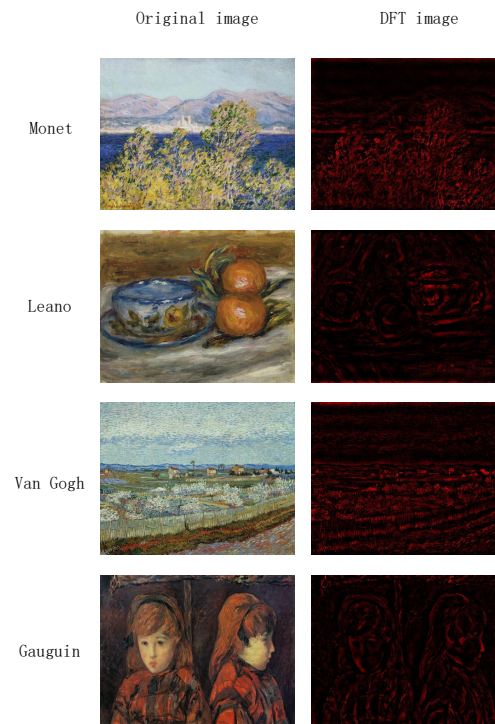


Figure 1. Fourier high-frequency channel filtering to extract brushstroke features.

3.3. Semantic Representation via State Space Modeling

Although Fourier decomposition effectively extracts physical brushstroke characteristics, frequency-domain representations still contain complex long-range spatial dependencies that cannot be adequately modeled through local convolutional operations.

To address this issue, the proposed framework employs the Mamba [21,24,25], State Space Model (SSM) [26] for semantic feature extraction from frequency-domain spectrograms.

Compared with conventional convolutional neural networks and Transformer architectures, Mamba exhibits linear computational complexity and an efficient selective state-space scanning mechanism, enabling the effective modeling of long sequential representations while maintaining computational efficiency.

By processing serialized frequency-domain feature maps, the Mamba encoder captures global stylistic dependencies distributed across distant frequency regions, thereby generating semantic representations that complement local brushstroke information extracted by Fourier decomposition [27,28].

The resulting semantic embedding [29] can, therefore, encode high-level stylistic organization beyond individual stroke textures, providing richer contextual information for artwork classification [30].

3.4. Dual-Branch Feature Injection Architecture

The left side shows the backbone network of the model, and the right side shows the specific implementation of the block. Its core is the Cross integration block, which adds the image and the image after the PFSD convolution block, then extracts features through convolution, and then performs mean pooling to achieve a uniform size. Finally, the four Cross integration are fused through the fusion block to perform feature fusion, as shown in Figure 2.

After feature disentanglement, the extracted physical-style representations and the original RGB image are simultaneously fed into a dual-branch classification architecture.

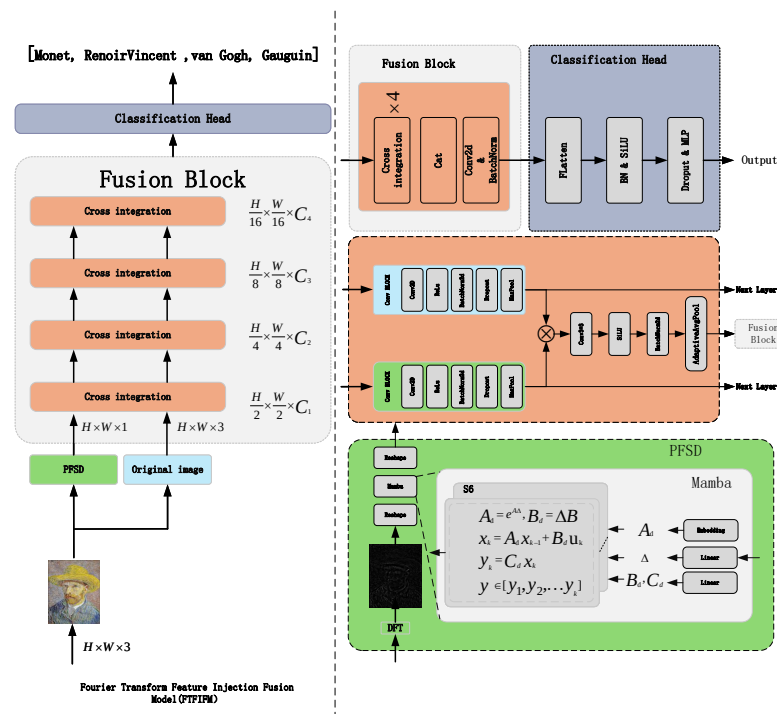


Figure 2. On the left is the Fourier Transform Feature Injection Fusion model architecture, and on the right are the individual modules and their internal implementations.

One branch processes the original image to preserve holistic semantic information, while the second branch processes the disentangled frequency-domain features to emphasize brushstroke and texture characteristics [31].

To minimize interference between heterogeneous feature spaces, the two branches are independently optimized throughout feature extraction. Their intermediate representations are subsequently integrated through a dedicated feature fusion module, which combines complementary evidence from multiple visual perspectives before final classification.

Compared with conventional feature concatenation strategies, the proposed disentangled feature injection mechanism explicitly preserves the independence of heterogeneous representations while enabling effective information integration at higher semantic levels. Consequently, the classifier is encouraged to focus on interpretable artistic characteristics [32] rather than learning entangled latent representations directly from raw pixels.

The overall architecture establishes a hierarchical computational pipeline consisting of physical features, stylistic representations, and semantic abstractions, providing an interpretable framework for AI-assisted artwork classification and human–AI collaborative visual analysis.

3.5. Design of Evaluation Metrics

In image classification tasks, to comprehensively evaluate model performance, this paper uses the following evaluation metrics:

$$\text{Top-1 Accuracy} = \frac{1}{N} \sum_{i=1}^N 1(\hat{y}_i^{(1)} = y_i)$$

$$\text{Precision} = \frac{TP}{TP + FP}$$

$$F1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

The above metrics are crucial for the comprehensive evaluation of image classification models. They have limitations when there is class imbalance—the model can achieve high accuracy by predicting all samples as the majority class, but it cannot identify the minority classes. In this case, Precision and the F1 score can more comprehensively assess the model's performance across each category: Precision reflects the reliability of the predictions, Recall reflects the model's ability to identify the target class, and F1 seeks a balance between the two. Top-1 accuracy, as a core metric for multi-class classification, directly measures the model's discrimination accuracy for the optimal class.

4. Result

4.1. Classification Performance of Deep Learning Models on Datasets of Four Impressionist Oil Painters

To verify our method's effectiveness, we selected three mainstream architectures—CNN, Transformer, and Mamba—as baseline controls, representing the core paradigms of convolution, attention mechanism, and state-space models, respectively.

To control for experimental variables, all models were trained with the same settings: a learning rate of 0.001, using an annealed cosine scheduler, an Adam optimizer, and 100 training epochs. The experiments used torch.float32 single-precision floating-point data. To prevent overfitting from affecting the comparison results, early stopping was employed during training; training was stopped if the accuracy did not improve after 20 epochs. Considering that Transformer and Mamba models require a large amount of data, and to control for experimental variables and eliminate errors caused by random weights, all models were pre-trained on the ImageNet-1K dataset, using pre-trained weights as initialization for fine-tuning. All data was finally tested on the test set.

In Table 4, the standard deviation reflects the fluctuation of predictions over each prediction run and is used to assess the model's operational stability.

We used the Pearson chi-square test to assess whether the differences between the comparison methods were statistically significant. This test was applied to a 2×2 contingency table containing observation frequencies (e.g., the number of correct predictions versus the number of incorrect predictions for the two models being compared). The null hypothesis was that there was no association between the models and the classification results (i.e., the two models had the same error distribution). The reported p -value corresponds to pairwise comparisons between the proposed model and each baseline model; a p -value < 0.05 indicates a rejection of the null hypothesis, suggesting a significant difference in performance. We applied Bonferroni correction to the multiple comparisons.

All confidence intervals were constructed using the bootstrap percentile method with 1000 resampling iterations. The confidence level was 95%. For a given metric, we generate bootstrap replicates by resampling the test set predictions with replacement, calculate the metric value for each replicate, and then use the 2.5th and 97.5th percentiles of the bootstrap distribution as the lower and upper bounds of the confidence interval.

The training results are shown in the table, as shown in Table 3. Standard deviation, significance, and confidence interval report, as shown in Table 4. Cross-validation, as shown in Table 5.

Table 3. Results.

Method	CNN-Based		
	TOP-1 ACC	Precision	F1
Monet	0.7818	0.7962	0.7890
Leano	0.8363	0.8363	0.8363
Van Gogh	0.8445	0.8260	0.8351
Gauguin	0.8223	0.8223	0.8222
Total	0.8212	0.8202	0.8206
Params	25M		

Table 3. *Cont.*

Method	Transformer-Based		
	TOP-1 ACC	Precision	F1
Monet	0.8181	0.7758	0.7964
Leano	0.8363	0.8214	0.8288
Van Gogh	0.8000	0.8163	0.8470
Gauguin	0.7777	0.7608	0.7692
Total	0.8080	0.8145	0.8103
Params	25M		
Method	Mamba-Based		
	TOP-1 ACC	Precision	F1
Monet	0.8181	0.8653	0.8411
Leano	0.8545	0.8703	0.8623
Van Gogh	0.8444	0.8085	0.8260
Gauguin	0.8222	0.8043	0.8131
Total	0.8340	0.8318	0.8337
Params	26M		
FTFIFM (Ours)			
	TOP-1 ACC	Precision	F1
Monet	0.8909	0.8750	0.8828
Leano	0.8546	0.8546	0.8545
Van Gogh	0.8000	0.8372	0.8181
Gauguin	0.8666	0.8478	0.8571
Total	0.8530	0.8536	0.8531
Params	10M		

Table 4. Standard deviation, significance, confidence.

	STD	<i>p</i>	Lower Confidence	Upper Confidence
CNN-based	0.0687	0.0478	0.8117	0.8309
Tranformer-based	0.0861	0.0445	0.7959	0.8200
Mamba-based	0.0927	0.0497	0.8211	0.8469
FTFIFM(ours)	0.0658	$p < 0.0001$	0.8237	0.8637

Table 5. The results are based on FTFIFM(ours) using a full dataset of 2000 images (stratified by artist) with 5-fold hierarchical cross-validation. We used scikit-learn to randomly partition the data into 5 disjoint subsets, each containing 4 training images and 1 validation image, and iterated 5 times. After each iteration, we report the performance metric for each fold and the average of the 5-fold cross-validation. The training environment is the same as described above.

Cross-Validation 5-Fold			
	TOP-1 ACC	Precision	F1
1-fold	0.8544	0.8432	0.8445
2-fold	0.8652	0.8613	0.8593
3-fold	0.8441	0.8528	0.8497
4-fold	0.8485	0.8453	0.8478
5-fold	0.8387	0.8413	0.8426
average	0.8502	0.8485	0.8488

It can be observed that our model outperforms other comparative models in overall performance, and particularly excels in the classification task of Monet's paintings (as shown in Confusion Matrix Figures 3 and 4). This is because Monet's works emphasize the expression of light, and such light features belong to high-frequency characteristics, belong to high-frequency characteristics that can be effectively extracted via high-pass filtering—as prior work in frequency-domain painting analysis has demonstrated [33,34], (as shown in Figures 5 and 6). This result further corroborates the role of feature injection in improving the performance of deep learning models, as well as its logical interpretability.

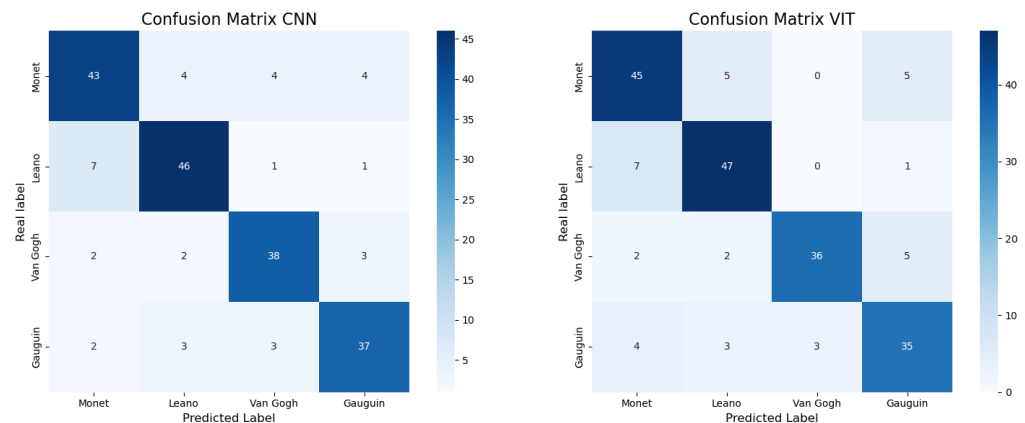


Figure 3. Illustration of CNN confusion matrix and VIT confusion matrix results.

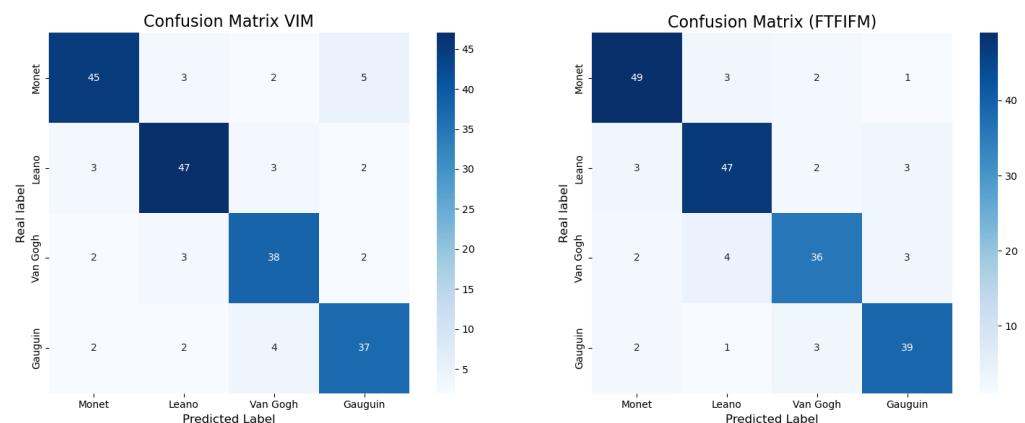


Figure 4. Illustration of VIM confusion matrix and FTFIFM (Ours) confusion matrix results.



Figure 5. Haystack and DFT image.

4.2. Ablation Experiment

To check how much each module we proposed contributes to the final performance, we set up an ablation experiment. The experiment was based on the baseline model, and we trained and evaluated model performance under different configurations by adding only one new module at a time. Specifically, we built the following variants one by one: a single-channel base model, a DFT Mamba single-channel model, a base Mamba dual-channel model, a base DFT dual-channel model, and the full base DFT Mamba model. All variants used the same training, validation, and test sets and were trained and evaluated with consistent hyperparameters and randomized weights. By comparing the differences in the main evaluation metrics across these variants, we could quantitatively analyze the individual contributions of each module and the interaction effects between modules. The results show that each proposed module brings a stable performance boost, and the full model achieved the best results, confirming the effectiveness of each component, as shown in Table 6.

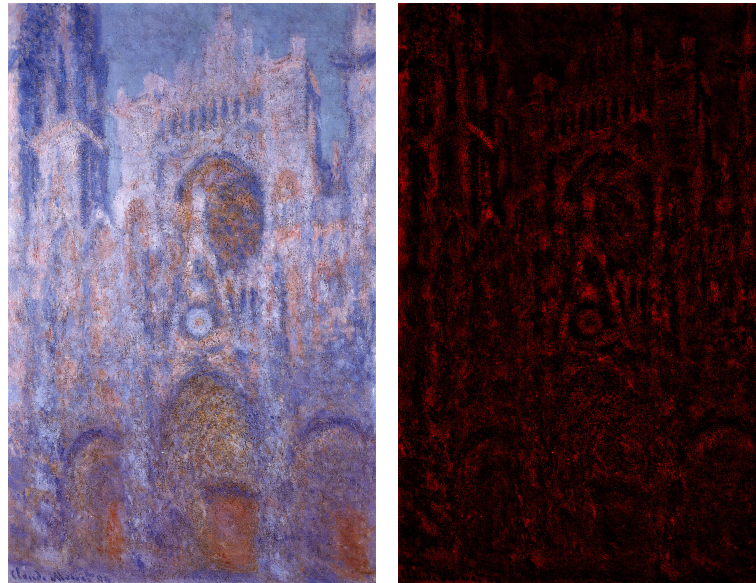


Figure 6. Rouen Cathedral and DFT image.

Table 6. Ablation experiment result.

Ablation Experiment			
Module Combination	TOP-1 ACC	Precision	F1
baseline (Single channel)	0.6350	0.6542	0.6444
DFT + Mamba (Single channel)	0.7653	0.7759	0.7705
baseline + Mamba (Dual channel)	0.6659	0.6941	0.6744
baseline + DFT (Dual channel)	0.7752	0.7879	0.7814
baseline + DFT + Mamba (ours)	0.8530	0.8536	0.8532

It can be observed that the pure basic CNN performs the worst, limited by its lightweight architecture and small number of parameters, making it difficult to extract features from complex images. Next is the two-channel combination of Baseline + Mamba. Due to the Mamba series' reliance on large-scale datasets, the Mamba branch, serving as a feature injector, does not provide a significant performance boost for such small-to-medium-scale classification tasks. In the single-channel combination of Mamba + DFT, DFT purifies the data, and Mamba, when working with the filtered data, sees an improvement in its state-space modeling capability, leading to better performance. After that comes the two-channel combination of Baseline + DFT, where DFT purifies the data as a feature injector and aligns data weights with the Baseline branch, further enhancing performance. Finally, the complete model (as shown in Figure 2)—consistent with our original design intention—reconstructs high-frequency features through the Mamba and DFT branches and fuses them with the features from the basic CNN, achieving an additional performance gain.

5. Discussion

5.1. Methodological Contributions and Limitations

The primary methodological contribution of this study lies in the introduction of a feature-guided input paradigm for artwork classification, which explicitly injects disentangled brushstroke representations into the visual recognition process. Rather than relying solely on end-to-end feature learning from raw images, the proposed framework incorporates both the original visual content and the disentangled physical and stylistic features derived through frequency-domain decomposition.

The necessity of this strategy stems from the observation that raw painting images contain substantial redundant visual information, including background illumination, large-scale color transitions, and compositional variations that may not contribute directly to artistic style discrimination. Conventional deep neural networks typically compensate for such redundancy by increasing model capacity, requiring a large number of parameters to implicitly learn discriminative representations from noisy inputs.

In contrast, the proposed feature injection mechanism explicitly provides brushstroke textures and semantic frequency representations that are closely associated with artistic style. By guiding the model toward perceptually meaningful design features, the classifier is encouraged to focus on discriminative subspaces that correspond to artistic intention rather than irrelevant visual variation. Consequently, the proposed lightweight architecture, containing approximately 10 million parameters, is able to outperform conventional baseline models with approximately 25 million parameters, demonstrating that explicit feature guidance can simultaneously improve classification performance and computational efficiency.

More importantly, the proposed framework reflects the design-oriented principle of feature disentanglement and guided recombination, transforming interpretable artistic knowledge into computational representations that can be directly incorporated into machine learning models. Such a paradigm provides a promising direction for developing explainable AI systems that support transparent human–AI collaboration rather than purely black-box prediction.

Nevertheless, several limitations should be acknowledged. First, the current framework focuses primarily on Impressionist paintings and has not yet been validated across a broader range of artistic movements. Second, the frequency-domain decomposition strategy mainly emphasizes brushstroke-related characteristics while other design elements, such as composition, spatial rhythm, and color harmony, remain insufficiently modeled. Finally, the current feature fusion mechanism adopts a relatively straightforward integration strategy, and more sophisticated evidence aggregation methods may further improve both classification accuracy and interpretability.

Comparability with Previous Work. It should be noted that this study cannot be directly quantitatively compared with published existing methods. First, most previous works utilize large-scale datasets with diverse style labels (e.g., WikiArt), while our dataset is specifically designed for learning brushstroke features from Impressionist artists; therefore, cross-dataset accuracy comparisons are unreliable. Second, the core innovation of our framework lies in interpretable decoupling and frequency-guided injection paradigms, rather than achieving state-of-the-art accuracy across a wide range of benchmarks; therefore, simply comparing accuracy does not fully reflect our contribution. Third, due to differences in backbone networks, pre-training strategies, and training protocols, we cannot fairly reproduce all previous methods without introducing implementation bias. Therefore, we conduct well-controlled internal comparisons under the same conditions against representative architectures (CNN, Transformer, Mamba), which strongly demonstrate the effectiveness of our method. We explicitly acknowledge these limitations of comparability and encourage readers to interpret our results within the established framework.

5.2. Implications for Design Research and Human–Computer Interaction

From the perspectives of design research and Human–Computer Interaction (HCI), the proposed feature injection framework offers important theoretical implications for AI-assisted creative systems.

5.2.1. Design Implications

The proposed approach facilitates a paradigm shift from conventional end-to-end black-box learning toward explicit representation of design intentions. Existing deep learning models typically entangle content and style within latent feature spaces, preventing designers from independently manipulating different dimensions of visual representation. Consequently, design knowledge remains implicitly encoded within model parameters and is difficult to interpret or control.

By explicitly disentangling brushstroke characteristics and semantic structural information through frequency-domain decomposition, the proposed framework transforms artistic knowledge into modular computational representations. Brushstroke features correspond to stylistic and textural expression, whereas semantic feature maps capture compositional and structural information. This modular representation enables individual design elements to be independently analyzed, manipulated, and recombined before being integrated into the final classification process.

Such a feature disentanglement-guided recombination mechanism closely aligns with the classical design methodology of analysis and synthesis, in which complex visual artifacts are first decomposed into constituent components before being recombined into coherent design solutions. Under this paradigm, designers can independently refine conceptual structures and expressive techniques while leveraging AI systems as collaborative creative partners rather than opaque computational tools.

Furthermore, by explicitly encoding prior design knowledge into the computational architecture, the proposed framework demonstrates that design theory can actively inform AI model construction rather than merely serving as an application domain for existing engineering methods.

5.2.2. Human–Computer Interaction Implications

From an HCI perspective, feature injection establishes a richer channel for design intention communication between humans and intelligent systems.

Traditional human–AI interaction in generative design primarily relies on textual prompts or sketch-based inputs, both of which struggle to precisely express subtle artistic characteristics such as brushstroke rhythm, texture organization, or localized semantic emphasis. Consequently, communication between designers and AI systems often remains ambiguous and indirect.

In contrast, the proposed framework enables users to directly manipulate interpretable feature representations. Designers may selectively modify, combine, or weight semantic feature maps and brushstroke vectors—for example, strengthening semantic responses in specific image regions or replacing stylistic feature sets derived from different paintings. Such operations transform latent representations into editable design prototypes, allowing creative intentions to be communicated through explicit computational features rather than natural language alone.

This capability fundamentally changes the role of humans within AI-assisted creative workflows. Instead of functioning merely as evaluators of generated outputs, users become active supervisors of the generative process, continuously guiding feature selection, representation learning, and decision formation. This interactive mechanism aligns with emerging paradigms of human–AI co-creativity, where intelligent systems augment rather than replace human creativity through collaborative exploration of design spaces.

Overall, the proposed feature disentanglement framework contributes not only to explainable artwork classification but also to the broader development of interpretable AI systems that support transparent interaction, controllable creativity, and collaborative visual understanding between humans and machines.

5.3. Toward Generalization: Adapting to Other Artistic Movements

While validated on Impressionism, extending our framework to other movements requires adaptive strategies, as each style exhibits distinct frequency-domain signatures.

Baroque relies on low-to-mid frequency cues—global contrast, dramatic chiaroscuro, and tonal gradation—rather than high-frequency brushstroke textures. Adaptation would require re-weighting filters toward band-pass or low-pass emphasis, with phase spectrum critical for encoding theatrical composition.

Cubism manifests as discontinuous spatial structures with sharp edges and fragmented planes, producing mid-to-high frequency components concentrated along geometric boundaries. Unlike Impressionism’s isotropic textures, Cubism demands phase-based structural disentanglement and multi-scale decomposition to capture local fragment geometries alongside global reorganization.

Abstract Expressionism splits into two subtypes: Action Painting exhibits highly anisotropic frequency distributions from directional, gestural strokes, requiring orientation-selective analysis; Color Field Painting yields extremely low-frequency-dominated spectra, demanding focus on large-scale chromatic organization and subtle gradients rather than texture [35].

These examples confirm that while our architectural principle is transferable, specific decomposition priorities—frequency band selection, phase vs. amplitude emphasis, and anisotropy modeling—must be adapted per movement, pointing toward a meta-framework for automatic style-specific optimization.

6. Conclusions and Future Work

This study presents an explainable AI framework for Impressionist painting classification that integrates feature disentanglement, frequency-domain analysis, and design-oriented representation learning. Departing from conventional end-to-end deep learning paradigms that encode raw pixels into entangled latent spaces, the proposed approach explicitly decomposes visual information into three complementary layers—physical brushstroke textures, stylistic frequency patterns, and semantic compositional structures—via Fourier-based filtering and state-space modeling. A dual-branch feature injection mechanism then fuses these disentangled representations with original image features, enabling a lightweight architecture (10M parameters) to outperform larger baseline models (25M parameters) while offering enhanced interpretability.

The framework makes three core contributions. First, it provides a mathematically rigorous and computationally efficient solution for decoupling artistic style from content in visual art, establishing a direct mapping between frequency-domain properties and human-interpretable design elements such as stroke granularity, directional energy, and texture organization. Second, by explicitly linking classification decisions to disentangled feature channels, it addresses the interpretability deficit of existing artwork classification systems, offering transparent reasoning pathways that can be inspected, questioned, and refined. Third, it repositions AI-based art analysis from a purely predictive task toward a human-AI collaborative paradigm, where users are empowered to actively guide model reasoning through feature-level interventions rather than passively consuming output labels.

From the perspective of design research and human-computer interaction, the proposed paradigm shifts the role of AI from that of an opaque oracle to that of an interpretable design partner. The disentangled representations serve as a shared vocabulary between human experts and computational systems, enabling a more nuanced communication of artistic intent, an iterative refinement of stylistic features, and an interactive exploration of visual semantics. This aligns with broader efforts in XAI and human-centered AI, where transparency, controllability, and user trust are prioritized alongside predictive performance.

Nevertheless, several limitations must be acknowledged. The current framework has been validated exclusively on Impressionist paintings, and its generalizability to other artistic movements with distinct stylistic signatures remains untested. Moreover, while brushstroke characteristics are effectively modeled through high-frequency components, other critical design dimensions—such as compositional rhythm, color harmony, spatial balance, and visual hierarchy—are not yet explicitly incorporated into the decomposition pipeline. The current feature fusion strategy, though effective, also represents a relatively straightforward integration mechanism that may not fully exploit the complementary nature of disentangled representations.

Future work will address these limitations along two primary directions. First, we plan to extend the framework to a broader range of artistic styles and historical periods, investigating how frequency-domain disentanglement can be adapted to capture movement-specific visual grammars beyond Impressionism. Second, building upon recent advances in multimodal large language models, we aim to develop a cross-modal feature-to-text explanation module that automatically translates disentangled feature maps—such as brushstroke orientation distributions or semantic attention patterns—into natural language descriptions accessible to non-expert audiences. This would bridge the gap between computational feature spaces and human artistic criticism, facilitating broader public engagement with AI-assisted art appreciation. Third, we will explore more sophisticated evidence aggregation strategies and richer multi-level feature decomposition schemes that integrate additional design semantics, including chromatic organization and compositional structure, into the disentanglement pipeline.

Ultimately, this work demonstrates that the convergence of feature disentanglement, frequency-domain analysis, and design-oriented AI offers a promising pathway toward next-generation computational aesthetics. By marrying mathematical rigor with human-centered interpretability, the proposed framework contributes to the development of transparent, collaborative, and creativity-enhancing AI systems that not only classify artworks but also deepen our understanding of artistic expression itself.

Author Contributions: Conceptualization, C.Y., G.S. and K.N.; methodology, C.Y., G.S. and K.N.; software, C.Y. and K.N.; writing—original draft, C.Y. and K.N.; writing—review and editing, Z.Q.; supervision, Z.Q.; project administration, Z.Q.; funding acquisition, Z.Q. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Key Project of Higher Education Research (2026) of Zhejiang Province Association for Higher Education (Project No. KT2026014), under the project “Research on the Reconstruction of AI-Integrated Art and Design Teaching Model under the Orientation of New-Quality Talent Cultivation.”

Data Availability Statement: The datasets presented in this article are not readily available because these images were acquired via web crawling and are bound by the terms of service of the source websites, which limit their use to academic research and forbid redistribution. Requests to access the datasets should be directed to the corresponding author upon reasonable academic request, and subject to permission from the copyright holders.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Rewald, J. *The History of Impressionism*; Museum of Modern Art Simon and Schuster: New York, NY, USA, 1973.
2. House, J. *Impressionism: Paint and Politics*; Yale University Press: New Haven, CT, USA, 2004.
3. Saleh, B.; Elgammal, A. Large-scale classification of fine-art paintings: Learning the right metric on the right feature. *arXiv* **2015**, arXiv:1505.00855.
4. Cetinic, E.; Lipic, T.; Grgic, S. Fine-tuning convolutional neural networks for fine art classification. *Expert Syst. Appl.* **2018**, *114*, 107–118. [[CrossRef](#)]

5. Koch, M.; Denzler, J.; Redies, C. Principal Component Analysis of Fourier Transforms Discriminates Visual Art from Other Image Categories. In *Perception*; Pion Ltd.: London, UK, 2009; Volume 38, p. 49.
6. Garcia, N.; Vogiatzis, G. How to read paintings: Semantic art understanding with multi-modal retrieval. In Proceedings of the European Conference on Computer Vision (ECCV) Workshops, Munich, Germany, 8–14 September 2018.
7. Castellano, G.; Vessio, G. Deep learning approaches to pattern extraction and recognition in paintings and drawings: An overview. *Neural Comput. Appl.* **2021**, *33*, 12263–12282. [\[CrossRef\]](#)
8. Karras, T.; Laine, S.; Aila, T. A style-based generator architecture for generative adversarial networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 16–20 June 2019*; IEEE: New York, NY, USA, 2019; pp. 4401–4410.
9. Li, Y.; Fang, C.; Yang, J.; Wang, Z.; Lu, X.; Yang, M.H. Universal style transfer via feature transforms. In *Proceedings of the Advances in Neural Information Processing Systems, Long Beach, CA, USA, 4–9 December 2017*; NIPS Foundation: San Diego, CA, USA, 2017; Volume 30.
10. Gatys, L.A.; Ecker, A.S.; Bethge, M. Image style transfer using convolutional neural networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016*; IEEE: New York, NY, USA, 2016; pp. 2414–2423.
11. Graham, D.J.; Field, D.J. Statistical regularities of art images and natural scenes: Spectra, sparseness and nonlinearities. *Spat. Vis.* **2007**, *21*, 149. [\[CrossRef\]](#)
12. Oppenheim, A.V.; Lim, J.S. The importance of phase in signals. *Proc. IEEE* **2005**, *69*, 529–541.
13. Pardo, A.; Pizzati, F.; Zhang, T.; Pondaven, A.; Torr, P.; Perez, J.C.; Ghanem, B. Matchdiffusion: Training-free generation of match-cuts. In *Proceedings of the IEEE/CVF International Conference on Computer Vision, Honolulu, HI, USA, 19–23 October 2025*; IEEE: New York, NY, USA, 2025; pp. 14973–14982.
14. Zhang, R.; Isola, P.; Efros, A.A.; Shechtman, E.; Wang, O. The unreasonable effectiveness of deep features as a perceptual metric. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018*; IEEE: New York, NY, USA, 2018; pp. 586–595.
15. Randen, T.; Husoy, J.H. Filtering for texture classification: A comparative study. *IEEE Trans. Pattern Anal. Mach. Intell.* **1999**, *21*, 291–310. [\[CrossRef\]](#)
16. Xu, B.; Yang, G. Interpretability research of deep learning: A literature survey. *Inf. Fusion* **2025**, *115*, 102721. [\[CrossRef\]](#)
17. Cooley, J.W.; Tukey, J.W. An algorithm for the machine calculation of complex Fourier series. *Math. Comput.* **1965**, *19*, 297–301. [\[CrossRef\]](#)
18. Andrews, H.C.; Pratt, W.K. Fourier transform coding of images. In Proceedings of the Hawaii International Conference on System Sciences (HICSS), Honolulu, HI, USA, 29–31 January 1968; pp. 677–679.
19. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016*; IEEE: New York, NY, USA, 2016; pp. 770–778.
20. Liu, Z.; Mao, H.; Wu, C.Y.; Feichtenhofer, C.; Darrell, T.; Xie, S. A convnet for the 2020s. In *Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), New Orleans, LA, USA, 21–24 June 2022*; IEEE: New York, NY, USA, 2022; pp. 11966–11976.
21. Zhu, L.; Liao, B.; Zhang, Q.; Wang, X.; Liu, W.; Wang, X. Vision mamba: Efficient visual representation learning with bidirectional state space model. *arXiv* **2024**, arXiv:2401.09417.
22. Wang, C.; Wu, H.; Jin, Z. Fourllie: Boosting low-light image enhancement by fourier frequency information. In *Proceedings of the 31st ACM International Conference on Multimedia, Ottawa, ON, Canada, 29 October 2023–3 November 2023*; ACM: New York, NY, USA, 2023; pp. 7459–7469.
23. Jiang, X.; Zhang, X.; Gao, N.; Deng, Y. When fast fourier transform meets transformer for image restoration. In *Proceedings of the European Conference on Computer Vision, Milan, Italy, 29 September–4 October 2024*; Springer Nature: Cham, Switzerland, 2024; pp. 381–402.
24. Gu, A.; Dao, T. Mamba: Linear-time sequence modeling with selective state spaces. *arXiv* **2023**, arXiv:2312.00752.
25. Liu, Y.; Tian, Y.; Zhao, Y.; Yu, H.; Xie, L.; Wang, Y.; Ye, Q.; Jiao, J.; Liu, Y. Vmamba: Visual state space model. *Adv. Neural Inf. Process. Syst.* **2024**, *37*, 103031–103063. [\[CrossRef\]](#)
26. Dao, T.; Fu, D.Y.; Saab, K.K.; Thomas, A.W.; Rudra, A.; Ré, C. Hungry hungry hippos: Towards language modeling with state space models. In Proceedings of the 11th International Conference on Learning Representations (ICLR), Kigali, Rwanda, 1–5 May 2023.
27. Huang, T.; Pei, X.; You, S.; Wang, F.; Qian, C.; Xu, C. Localmamba: Visual state space model with windowed selective scan. In *Proceedings of the European Conference on Computer Vision, Milan, Italy, 29 September–4 October 2024*; Springer Nature: Cham, Switzerland, 2024; pp. 12–22.
28. Pei, X.; Huang, T.; Xu, C. Efficientvmamba: Atrous selective scan for light weight visual mamba. In *Proceedings of the AAAI Conference on Artificial Intelligence, Philadelphia, PA, USA, 25 February–4 March 2025*; AAAI: Washington, DC, USA, 2025; Volume 39, pp. 6443–6451.

29. Zheng, K.; Zhang, Y.; Wu, W.; Lu, F.; Ma, S.; Jin, X.; Chen, W.; Shen, Y. Dreamlip: Language-image pre-training with long captions. In *Proceedings of the European Conference on Computer Vision, Milan, Italy, 29 September–4 October 2024*; Springer Nature: Cham, Switzerland, 2024; pp. 73–90.
30. Ma, K.; Lee, S.; Ma, X.; Chen, H. Fine art image classification and design methods integrating lightweight deep learning. *Sci. Rep.* **2025**, *15*, 33006. [[CrossRef](#)] [[PubMed](#)]
31. Panda, R.; Fein, D.; Singhal, A.; Fiore, M.; Agrawala, M.; Bohacek, M. LouvreSAE: Sparse Autoencoders for Interpretable and Controllable Style Transfer. *arXiv* **2025**, arXiv:2512.18930.
32. Dang, Y.; Duan, F.; Chen, J. Oil-Painting Style Classification Using ResNet with Conditional Information Bottleneck Regularization. *Entropy* **2025**, *27*, 677. [[CrossRef](#)] [[PubMed](#)]
33. Wang, H. Research on classification method of painting features based on stochastic forest algorithm. *Int. J. Inf. Commun. Technol.* **2023**, *23*, 91–105. [[CrossRef](#)]
34. Bai, Y.; Guo, Y.; Wei, J.; Lu, L.; Wang, R.; Wang, Y. Fake generated painting detection via frequency analysis. In *Proceedings of the IEEE International Conference on Image Processing (ICIP), Online, 25–28 October 2020*; IEEE: New York, NY, USA, 2020; pp. 1256–1260.
35. Singh, B.; Chaudhary, A.; Devipriya, A. Frequency-Aware Visual Restoration and Cross-Stream Multimodal Attention for Action Classification in Degraded Videos. *IEEE Access* **2026**, *14*, 89062–89082. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.