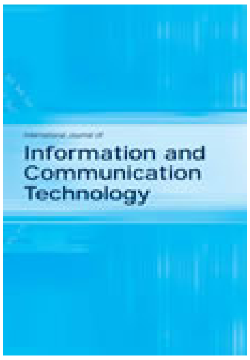


International Journal of Information and Communication Technology

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

Research on fine-grained classification algorithm of oil painting schools based on multilevel SVM and feature engineering

Qi Xie, Chaobin Wang

DOI: 10.1504/IJICT.2026.10075530

Article History:

Received:	29 September 2025
Last revised:	12 November 2025
Accepted:	16 November 2025
Published online:	22 January 2026

Research on fine-grained classification algorithm of oil painting schools based on multilevel SVM and feature engineering

Qi Xie* and Chaobin Wang

Academy of Painting,
Hubei Institute of Fine Arts,
Wuhan, Hubei, 430205, China
Email: kaixinheihei111@163.com
Email: xiexiemeimei123@163.com
*Corresponding author

Abstract: This paper aims to address these issues by proposing a fine-grained classification algorithm system that integrates multi-level support vector machines (SVMs) and feature engineering. This system combines data pre-processing, feature enhancement, and multi-level SVM classification to construct a hierarchical decision-making structure of genre clusters → genres → periods. Experiments show that the algorithm achieves an accuracy of 92.3% on a diverse dataset, with a macro F1 score of 0.915. Furthermore, in robustness tests, the accuracy only drops to 88.0% under noise perturbation, 87.3% under blur perturbation, and 85.5% under occlusion perturbation. The cross-dataset generalisation accuracy reaches 85.2%. External validation indicates an average accuracy of 82.3% for non-Western oil paintings, while also demonstrating high interpretability. This paper contributes an innovative technical approach for fine-grained classification of oil painting genres; it enhances accuracy, robustness, and interpretability, and lays the foundation for subsequent lightweight design, multimodal fusion, and cross-media expansion research.

Keywords: multi-level SVM; support vector machines; feature engineering; oil painting; genre; classification.

Reference to this paper should be made as follows: Xie, Q. and Wang, C. (2026) 'Research on fine-grained classification algorithm of oil painting schools based on multilevel SVM and feature engineering', *Int. J. Information and Communication Technology*, Vol. 27, No. 1, pp.77–99.

Biographical notes: Qi Xie obtained his Bachelor's degree from the Department of Oil Painting at the Hubei Institute of Fine Arts, and subsequently earned a Master's degree from the same department. He is currently teaching in the Department of Oil Painting at the School of Painting, Hubei Institute of Fine Arts. His primary professional fields include oil painting art research and contemporary art studies, with a current research focus on digital painting.

Chaobin Wang is a Professor and Master's Supervisor in the Oil Painting Department of the Oil Painting College, Hubei Institute of Fine Arts. The works have participated in important exhibitions such as the 'First Guangzhou Art Biennale', the 'Second National Oil Painting Exhibition', the 'Third China Oil Painting Annual Exhibition', and the 'Outstanding Award of Hubei Provincial Art Exhibition'.

1 Introduction

As an important topic in the intersection of computer vision and artistic digitalisation, the identification of oil painting genres is attracting increasing attention all over the world. With the acceleration of the digitalisation process of museums, galleries and private collections, how to efficiently and accurately classify massive oil paintings by genres has become the core demand of art heritage protection, digital collection management and even art market analysis. The traditional classification method is highly dependent on the manual identification of art history experts, which is not only subjective and inefficient, but also difficult to apply on a large scale. Especially, in the face of cross-cultural and cross-period diversified artistic styles, its generalisation ability shows obvious limitations. In recent years, image classification methods based on deep learning have made breakthroughs in natural image processing and shown certain potential in the field of art (Taori et al., 2020). However, for the fine-grained classification task of oil painting genres, existing technologies still face three core challenges. First, the differences in visual features between different genres are extremely subtle. For example, Impressionism and Pointillism are highly similar in brushstroke techniques and colour application, and it is difficult for traditional convolutional neural networks (CNNs) to capture these subtle differences (Zeng, 2022). Secondly, there are significant times and regional evolutions within the same genre. For example, Baroque style shows different visual characteristics in different regions such as Italy and the Netherlands, resulting in greater differences within classes than inter-class differences (Chen and Yang, 2021). Third, high-quality labelled data in the art field is relatively scarce, and its distribution is extremely uneven. Therefore, deep learning methods are easy to fall into overfitting, and their generalisation ability is limited (Hu and Yang, 2024).

These limitations severely restrict the accuracy and reliability of digital appraisal of oil painting art (Guo, 2022). To solve the above problems in the system, this paper proposes a fine-grained classification algorithm that integrates multi-level support vector machines (SVMs) and deep feature engineering. By combining the feature extraction ability of deep CNNs with the discriminative ability of multi-level SVM classifiers, a hierarchical classification decision mechanism is constructed.

The purpose of this study is to address the three core challenges faced in fine-grained classification of oil painting genres: subtle differences in visual features between genres, spatiotemporal style evolution within the same genre, and scarce annotation data. By proposing a hierarchical classification algorithm that integrates multi-level SVMs and deep feature engineering, we innovatively embed prior knowledge in the art field into a machine learning framework; In terms of research methods, a classification paradigm that combines feature discriminative power and structural adaptability has been constructed with feature enhancement (LAB colour space contrast optimisation and contrast limited adaptive histogram equalisation (CLAHE) texture enhancement), deep feature refinement [VGG-16+SEBlock (squeeze-and-excitation block) channel attention mechanism], and multi-level SVM classification (school cluster → school → period tree decision structure) as the core. Its innovation is reflected in the combination of traditional feature engineering and modern deep learning advantages, and for the first time, a hierarchical decision mechanism is introduced in art classification to mimic expert discrimination logic; The contribution of this article is not only reflected in the significant improvement of classification performance (accuracy of 92.3%, an increase of over 3.6% compared to the baseline model), but also in providing an interpretable and robust technical path for cross-cultural and cross period art image analysis, which combines theoretical innovation value and engineering practical significance.

2 Related work

As an interdisciplinary field between computer vision and digital humanities, automatic classification of oil painting genres has received widespread attention in recent years. Existing research has explored from multiple perspectives such as feature extraction, classification model design, and cross-domain transfer, but there are still significant challenges in fine-grained classification, scarce annotation, and style evolution modelling. This section systematically reviews and analyses existing literature around the themes of feature extraction, classification methods, and cross-domain applications.

2.1 Handcrafted features and traditional machine learning methods

Early research mainly relied on manually designed visual features and traditional classifiers. Huang (2021) proposed an adaptive texture feature extraction method based on reaction-diffusion equations, which enhances the representation ability of strokes and textures by simulating the physical diffusion process, achieving an accuracy rate of about 78% on a smaller-scale dataset. However, this method lacks sufficient utilisation of colour and global composition information, and has limited feature expression ability. Nunez-Garcia et al. (2022) combined colour histograms and local binary pattern (LBP) texture features, using random forests for genre classification, achieving a macro F1 score of 0.81 on multiple style datasets. Its shortcoming lies in the relatively simple feature fusion strategy, which fails to effectively capture high-level semantic information.

Although these methods are highly interpretable and have low computational overhead, they rely heavily on domain knowledge and have limited feature representation capabilities, making it difficult to handle subtle differences between genres and intra-class diversity issues.

2.2 Deep learning and end-to-end classification model

With the development of deep learning, methods based on CNNs have gradually become mainstream. Zhang (2024) proposed the ResNet-NTS (non-local tensor sampling) network, which captures long-distance spatial dependencies through a non-local tensor sampling module, achieving an accuracy of 88.7% on the WikiArt dataset. This model still has room for improvement in terms of computational complexity and training stability. Cheng et al. (2024) adopted multi-feature fusion and style transfer techniques, combined with VGG and attention mechanisms to identify painting styles and emotional tendencies, achieving a macro F1 score of 0.872 in the joint task, but their multi-task framework had optimisation conflict issues.

On the other hand, Azimi et al. (2024) proposed a local feature learning method based on generative adversarial networks (GAN) and stroke analysis for the task of oil painting forgery detection. This method achieves forgery area localisation through patch-level classification, achieving an accuracy rate of over 90%. However, this method requires high-resolution input and has extremely high requirements for annotation quality. Cao et al. (2024) proposed a classification method that integrates image enhancement and

CNN for underwater oil spill recognition. Although this method is robust in specific environments, its adaptability to artistic images is poor.

Deep learning methods have significantly improved feature representation capabilities, but they are still limited by data scarcity and interference from similarity between categories, and there are deficiencies in computational cost and interpretability.

2.3 Cross-domain application and multimodal extension

Some studies have explored the integration of art classification with other fields. Wang et al. (2022) developed an intelligent painting recognition system based on image perception for multimedia enterprise applications, integrating colour, texture, and layout features. The system achieved an accuracy rate of 83.5% in real-world testing in an enterprise environment, but its generalisation ability was relatively weak. Li et al. (2022) constructed a visual communication effect evaluation model for art works based on machine learning, using regression analysis to quantify aesthetic indicators. The Pearson correlation coefficient reached 0.79 on a cross-cultural dataset, but the feature design was not closely integrated with art theory.

In terms of educational applications, Chiu et al. (2024) designed an art education assistance system based on deep learning to enhance students' appreciation ability through style analysis and creation guidance. The experiment showed that the quality of students' works improved significantly ($p < 0.01$), but the model did not consider individual aesthetic differences. In addition, Cao et al. (2021) applied an improved Inception-v3 model to classify ancient mural dynasties, achieving an accuracy rate of 85.3%, providing a feasible path for historical artefact analysis, but did not solve the problem of continuous style evolution.

These studies demonstrate the breadth of applications of computational art, but most methods rely on general-purpose network architectures and lack targeted designs for specific challenges in the art field.

2.4 Feature engineering and optimisation strategies

Feature optimisation is a key step to improve classification performance. Bi et al. (2020) proposed a feature learning method combining genetic programming and image descriptors to adaptively generate discriminative feature subsets, which improved the accuracy by about 5% compared to traditional methods on multiple standard datasets, but the algorithm complexity was high. Yangxiaoxiao (2021) used AI technology for image modal analysis and recognition, enhanced feature interpretability through tensor decomposition dimensionality reduction and sparse coding, and achieved an accuracy of 82.3% on the art design dataset, but it was sensitive to noise.

In the field of cultural heritage protection, Wang et al. (2024) developed a processing flow based on 3D-CNN and hyperspectral imaging for the virtual restoration of mould-damaged murals, achieving an 18.7% improvement in restoration accuracy compared to traditional methods, but relying on specialised acquisition equipment. Zeng et al. (2020) combined CNN and nearest neighbour algorithms to achieve controllable digital restoration of ancient paintings. The work published in pattern recognition Letters emphasised the balance between user interaction and algorithm output, but there are still deficiencies in terms of automation.

Although feature engineering research has promoted technological development from different perspectives, most methods have not systematically considered the hierarchical characteristics of artistic images and expert prior knowledge.

Table 1 summarises the aforementioned research in terms of model methodology, key findings, and limitations

Table 1 Summary of existing research

Research model	*Result*	*Deficiency*
Reaction diffusion feature extraction	Accuracy is about 78%	Insufficient utilisation of colour and global features
Colour and texture feature fusion	Macro F1 score 0.81	The feature fusion strategy is simple
ResNet-NTS	Accuracy rate: 88.7%	High computational complexity
Multi-feature fusion and style transfer	Macro F1 score 0.872	Multi-task optimisation conflict
GAN brush analysis	Accuracy > 90%	Rely on high-resolution input
Image enhancement + CNN	Strong robustness in specific environment	Poor adaptability of artistic images
Enterprise intelligent recognition system	Accuracy rate of 83.5%	Weak generalisation ability
Visual communication evaluation model	Correlation coefficient 0.79	Weak integration with art theory
Feature learning in genetic programming	Accuracy improvement of 5%	High algorithm complexity
Tensor decomposition and sparse coding	Accuracy rate of 82.3%	sensitive to noise
3D-CNN hyperspectral processing	The precision of the repair has been enhanced by 18.7%	Rely on professional equipment
CNN and nearest neighbour repair	Good user interaction	Inadequate automation

In summary, existing research has made significant progress in feature design, deep network optimisation, and application expansion, but still generally faces problems such as insufficient fine-grained discrimination ability, strong dependence on labelled data, limited cross-domain generalisation performance, and weak interpretability. Especially in scenarios where there is high similarity between genres, significant intra-class variation, and scarce data, current methods have not yet achieved a balance between accuracy, robustness, and practicality.

To this end, this paper proposes a classification algorithm that integrates multi-level SVM and deep feature engineering. Through LAB colour space and CLAHE texture enhancement, VGG-16 and SEBlock attention feature extraction, and hierarchical SVM classification structure, the system addresses the above challenges, enhancing model interpretability and cross-dataset generalisation ability while ensuring high-precision classification.

3 Classification algorithm model

The identification of oil painting genres is an important subject in the cross field of computer vision and artistic digitalisation. Traditional genre classification methods mostly rely on the subjective judgment of art experts, and it is difficult to realise the rapid and accurate classification of large-scale paintings (Kleynhans et al., 2020). In recent years, image classification methods based on deep learning have shown great potential in the field of art. However, there are still the following challenges for fine-grained oil painting genre classification tasks.

1 There are subtle differences in visual features between different genres, such as the high similarity between Impressionism and Pointillism in brushstrokes and colour application.

2 There is style evolution within the same school, such as Baroque style shows obvious differences in different periods and regions.

3 The available labelled data is relatively limited, which restricts the generalisation ability of deep learning models.

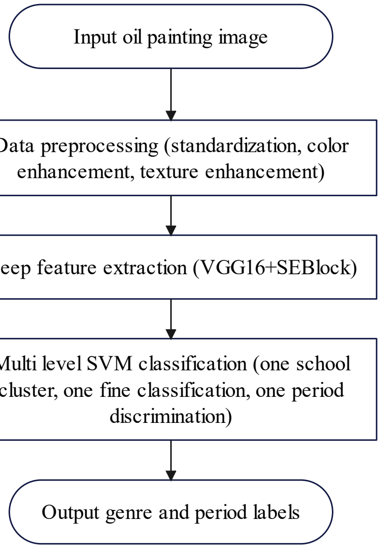
To address the above challenges, this section proposes a fine-grained classification algorithm for oil painting genres based on multi-level SVM and depth feature engineering. This algorithm combines the feature extraction ability of deep CNN with the discriminant ability of multi-level SVM classifier, and achieves accurate identification of oil painting genres through feature enhancement and hierarchical classification strategies. This method not only performs multi-dimensional optimisation at the feature representation level, but also introduces hierarchical decision-making mechanism in the classification structure, which significantly improves the accuracy and robustness of fine-grained classification.

3.1 Overall algorithm framework

The flow of the fine-grained classification algorithm of oil painting genres proposed in this paper is shown in Figure 1, which includes three core modules: data pre-processing and feature enhancement module, deep feature extraction module, and multi-level SVM classification module. The whole framework gradually refines and utilises image features in a hierarchical and multi-stage way, and finally realises the progressive recognition of oil paintings from style clusters, specific genres to creative periods.

The design of the frame fully considers the characteristics of artistic images. On the one hand, oil paintings have a high degree of complexity and diversity in colour, texture and composition; on the other hand, the distinction between different genres often relies on detailed characteristics. Therefore, the algorithm emphasises the enhancement and purification of original image information in the early stage, and focuses on the extraction and structured classification of high-level semantic features in the middle and late stages.

Figure 1 Flow of fine-grained classification algorithm of oil painting genres



3.1.1 Data pre-processing and feature enhancement

Data pre-processing is a key step to improve model generalisation capabilities. According to the characteristics of oil painting images, this algorithm designs a pre-processing process including three sub-modules: image standardisation, colour space enhancement and texture feature enhancement.

1 Image normalisation: all input images are adjusted to a uniform size (224×224 pixels) and normalised. Specifically, the input image I is transformed as follows (Chen, 2022):

$$I_{norm} \frac{I-\mu}{\sigma} \tag{1}$$

Among them, μ and σ are the mean and standard deviation of the image set, respectively. This step not only unifies the input scale but also reduces the noise caused by differences in lighting and acquisition conditions.

2 Oil painting attaches great importance to colour expression, so this paper chooses to enhance contrast in LAB colour space. The LAB colour model can better approximate human visual perception, where the L channel represents lightness and the A and B channels represent green-red and yellow-blue opposite colours, respectively. The enhancement formula is as follows:

$$L^* = \alpha \cdot l, \alpha^* = \beta \cdot a, b^* = \gamma \cdot b \tag{2}$$

Among them, α, β and γ are adjustment coefficients, which optimise the brightness, green-red and yellow-blue channels respectively. This enhancement strategy can highlight the unique colour application style of paintings and provide more distinctive colour information for subsequent feature extraction.

3 Enhancement of texture features: the brushstrokes and texture of oil paintings are important visual clues to distinguish genres. In this paper, adaptive histogram equalisation (CLAHE) is used to enhance stroke texture details:

$$L_{enhanced} = CLAHE\left(I, C_{limit}, T_{size}\right) \tag{3}$$

Among them, C_{limit} is the contrast limit and T_{size} is the tile size. By enhancing local contrast, we can highlight the brushstroke details and material properties of the painting without amplifying global noise.

3.1.2 Deep feature extraction module

This paper constructs a deep feature extraction network based on VGG-16 architecture, which has the following advantages.

1 The hierarchical structure can capture visual information from low-level edges, textures to high-level semantic features.

2 The general features learned by the pre-trained model on ImageNet provide good parameter initialisation for artistic images.

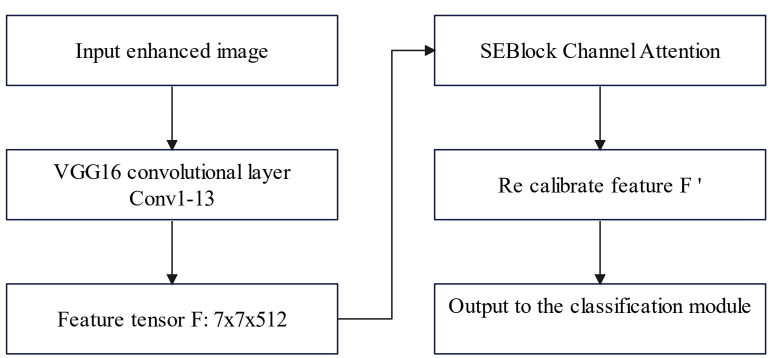
3 The medium-depth network architecture is suitable for transfer learning to the art field with relatively limited data.

The feature extraction process can be expressed as (Jiang and Yang, 2022):

$$F = \Phi\left(I_{enhanced}; \theta\right) \tag{3}$$

Among them, $\Phi(\cdot)$ represents the feature extraction function composed of the first 13 convolutional layers of the VGG16 network, θ is the network parameter, and $F \in \mathbb{R}^{7\times7\times512}$ is the extracted deep feature tensor.

Figure 2 Deep feature extraction process



In order to enhance feature discrimination, this paper introduces a feature refining mechanism-compression-excitation module (SEBlock). The module generates channel-level statistics through global average pooling, then learns the importance weight of each channel through two fully connected layers, and finally recalibrates the original features:

$$F_{refiend} = SEBlock(F) \otimes F \tag{4}$$

Among them, *SEBlock* is the compression-excitation module, which enhances the response of important channels through feature recalibration.

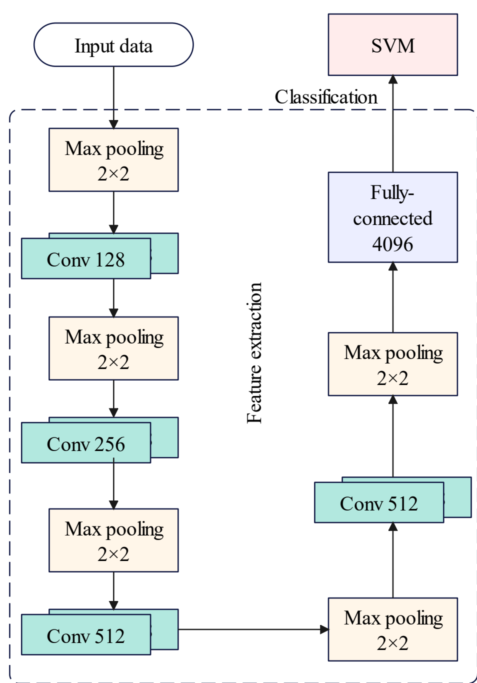
The flow chart of Figure 2 further illustrates the specific steps of depth feature extraction.

The process includes four sub-steps: convolutional feature extraction, spatial information aggregation, channel attention weighting, and global feature embedding. Finally, a highly discriminant feature vector is output to provide robust feature representation for subsequent classification.

3.1.3 Multi-level SVM classification module

Aiming at the hierarchical characteristics of oil painting genres, this paper designs a multi-level SVM classification structure (as shown in Figure 3). The structure consists of three levels: genre cluster classification, genre sub-classification and period identification, which imitates the identification process of art experts first identifying general style categories, then distinguishing specific genres, and finally judging the creative period.

Figure 3 Multi-level SVM classification structure (see online version for colours)



- Level 1: Genre cluster classification

  The genres with similar styles are clustered into superclasses, (e.g., Impressionism, Pointillism, and Post-impressionism are clustered into the 'Modernism' superclass), and the RBF kernel SVM is used:

$$f_1(x) = sign\left(\sum_{i=1}^{N} \alpha_i y_i K(x_i, x) + b\right) \tag{5}$$

Among them, $K(x_i, x) = \exp(-\gamma\|x_i - x\|^2)$ is the RBF kernel function. RBF kernel can handle nonlinear separable problems and is suitable for separating different genre clusters in feature space.

- Level 2: Genre subclassification

  Fine-grained genre discrimination is performed within each superclass, and the linear SVM is used:

$$f_2(x) = w^T x + b \tag{6}$$

  Linear SVM performs well in high-dimensional feature space, and has high computational efficiency, so it is suitable for dealing with subtle differences between genres.

- Level 3: Period identification

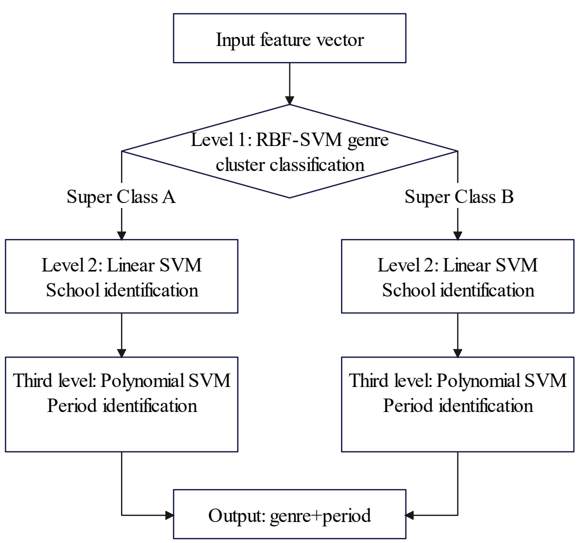
  To further distinguish the creative period of specific genres (such as the early Renaissance and the high Renaissance), the polynomial kernel SVM is used:

$$f_3(x) = \left(w^T \varphi(x) + b\right)^d \tag{7}$$

Among them, d is the polynomial order, and $\varphi(\cdot)$ is the feature mapping function. Polynomial kernels can capture high-order correlations between features and are suitable for handling gradual style changes across different periods within the same genre.

The flowchart shown in Figure 4 shows the multi-level SVM classification decision process.

Figure 4 Multi-level SVM classification decision process



The process adopts a tree-like decision structure, and the classification results of each level determine the classifier used in the next level, which not only improves the classification accuracy, but also reduces unnecessary computational overhead.

3.2 Feature optimisation and selection strategy

3.2.1 Deep characteristic distillation

In order to solve the problems of high feature dimension and high computational complexity, this paper uses knowledge distillation technology to compress feature dimension while retaining discriminant information. Using a lightweight student network to learn the feature representation of the teacher network (the original feature extraction network), the distillation loss function is:

$$L_{distill} = \lambda L_{KL}\left(p_{teacher}, p_{student}\right) + (1+\lambda) L_{CE}\left(y, p_{student}\right) \tag{8}$$

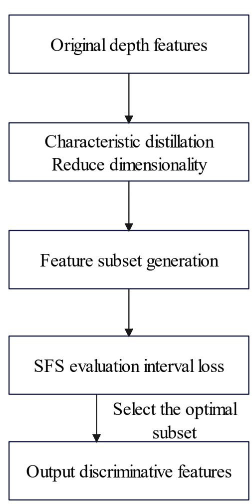
Among them, L_{KL} is the KL divergence loss, which measures the difference in the output distribution of the teacher network and the student network, L_{CE} is the cross entropy loss, which ensures that the student network maintains the classification performance, and λ is the balance parameter.

3.2.2 Discriminant feature selection

The sequence forward selection (SFS) algorithm is used to screen the most discriminant feature subset. SFS is a greedy search algorithm that starts with an empty set and adds one feature at a time that best improves classification performance until the performance is no longer significantly improved. The evaluation function of the feature subset S is:

$$S^* = \arg\max_{S \subseteq F} J(S) \tag{9}$$

Figure 5 Feature optimisation and selection process



Among them, $J(S)$ is the evaluation function for the feature subset S, and the margin loss is used as the evaluation criterion. This evaluation criterion is directly related to the margin of the classifier, ensuring that the selected feature subset has the strongest discriminative ability.

The feature optimisation and selection process is shown in Figure 5.

The process includes four stages: feature importance assessment, iterative feature subset search, verification set performance assessment and final feature subset determination, ensuring that classification performance is maintained or even improved while reducing dimensions.

3.3 Model optimisation and training strategy

3.3.1 Loss function design

Combined with the multi-task learning framework, the overall loss function is:

$$L_{total} = \sum_{k=1}^{3} \lambda_k L_k + \eta \|\Theta\|_2^2 \tag{10}$$

Among them, L_k is the hinge loss of the k^{th} level SVM, λ_k is the weight coefficient, and η is the regularisation parameter. The multi-task learning framework enables classifiers at all levels to share feature representations and promote each other, improving overall performance.

3.3.2 Parameter optimisation

Particle swarm optimisation (PSO) algorithm is used to optimise the SVM parameters. PSO is an optimisation algorithm based on swarm intelligence, which finds the optimal solution by simulating the foraging behaviour of birds. Each particle represents a candidate solution (a set of SVM parameters) whose position and velocity are updated as:

$$v_i^{t+1} = \omega v_i^t + c_1 r_1 \left(p_i^t - x_i^t\right) + c_2 r_2 \left(g^t - x_i^t\right) \tag{11}$$

$$x_i^{t+1} = x_i^t + v_i^{t+1} \tag{12}$$

Among them, v_i and x_i represent the particle's velocity and position, respectively, ω is the inertia weight, c_1 and c_2 are learning factors, and r_1 and r_2 are random numbers. PSO can efficiently search high-dimensional parameter spaces to find the globally optimal or near-optimal SVM parameter configuration.

3.4 Algorithm complexity analysis

If the number of training samples is N, the feature dimension is D, and the number of SVMs is M, the time complexity of the algorithm is:

$$O(M \cdot N^2 \cdot D) \tag{13}$$

By introducing feature selection and model compression techniques, the actual complexity can be reduced to:

$$O(M \cdot N \cdot \log N \cdot D') \tag{14}$$

Among them, $D' \ll D$ is the selected feature dimension.

The classification algorithm proposed in this paper, which combines multi-level SVM with feature engineering, effectively solves the challenges in fine-grained classification of oil paintings through carefully designed feature enhancement, deep feature extraction, hierarchical classification and optimisation strategies, and provides effective technical support for digital identification of artistic works.

4 Test

4.1 Test methods

The purpose of this test is to systematically verify the comprehensive performance of the proposed fine-grained classification algorithm of oil painting genres based on multi-level SVM and feature engineering in terms of accuracy, robustness, practicality and interpretability. Through comparative analysis with current advanced baseline models, ablation experiments and multi-dimensional evaluation, this paper proves its effectiveness and superiority in solving fine-grained classification challenges of oil painting genres (such as subtle differences in features between genres, style evolution within the same genre, and generalisation ability under limited labelled data).

This experiment uses the following global publicly available oil painting datasets:

1 WikiArt dataset (https://www.wikiart.org/), containing over 80,000 oil paintings covering from the 13th century to modern times, labelled with genre, period, artist, and style tags, available through its official API or Kaggle platform (https://www.kaggle.com/datasets/ipythonx/wikiart-gangogh) obtain.

2 Metropolitan Museum of Art Open Access dataset (https://www.metmuseum.org/about-the-met/policies-and-documents/open-access), including high-resolution images and metadata of approximately 30,000 artworks.

3 Rijksmuseum dataset (https://www.rijksmuseum.nl/en/research/data) provide digital resources for over 100,000 artworks.

4 The National Gallery (London) dataset (https://www.nationalgallery.org.uk/about-us/policies-and-documents/image-licensing), containing over 20,000 Western oil paintings, enhances coverage from the Renaissance to modern styles.

5 The Asia Art Archive dataset (https://www.aaa.org.hk/en/collection), featuring over 15,000 modern and contemporary Asian oil paintings (such as works from China, Japan, and India), fills the gap in non-Western art.

6 The Smithsonian American Art Museum dataset (https://americanart.si.edu/open-access), containing 10,000 American oil paintings, spans from the colonial period to contemporary genres, enhancing regional diversity.

All datasets were subjected to a unified pre-processing method: first, image normalisation was performed (the input image was uniformly adjusted to 224 × 224 pixels, and normalisation were applied based on the mean (μ = 0.485, 0.456, 0.406) and standard deviation (σ = 0.229, 0.224, 0.225) calculated from the entire training set). Then, contrast enhancement was performed in the LAB colour space (adjustment coefficients α = 1.2, β = 1.1, γ = 1.1). Finally, adaptive histogram equalisation (CLAHE, parameters C_{limit} = 2.0, Tsize = 8 × 8) was used to enhance the texture details of the strokes, forming an enhanced texture. The feature table is presented to support subsequent deep feature extraction and classification.

The experimental subjects include the multi-level SVM model proposed in this article (experimental group) and currently advanced baseline models (control group), such as ResNet-50, EfficientNet-B4, ViT Base, and traditional SVM and CNN combined models. The experimental grouping is divided into performance testing (comparing accuracy, F1 score, inference time), robustness testing [adding Gaussian noise (σ = 0.05), Gaussian blur (kernel size = 5), and random occlusion (20% area) disturbance], practicality testing (cross dataset generalisation and computational efficiency evaluation), ablation testing (gradually removing feature enhancement, multi-level classification or feature selection modules), and interpretability testing (feature visualisation and attention analysis). In addition, designers participated in the evaluation experiment and invited 30 art experts (including ten art history professors, 15 museum researchers, and five professional art appraisers) to independently annotate 500 randomly selected paintings, comparing the consistency (Kappa coefficient) between the model and human experts. All experiments were repeated three times using 5-fold cross validation, with training, validation, and testing sets divided into 70%, 15%, and 15% ratios. The hyperparameters were uniformly optimised through PSO to ensure fairness.

4.2 Results

Based on the original experiments, this section introduces stratified cross-validation (stratified 5-fold cross-validation repeated five times), external validation set (20% of the works extracted from the newly added dataset as an independent test set), data augmentation validation (style transfer enhancement based on AdaIN network), and bias analysis (quantification of regional cultural bias) to comprehensively enhance the statistical significance of the results. All experiments are based on the expanded diversified dataset (including WikiArt, Metropolitan Museum, Rijksmuseum, The National Gallery, Asia Art Archive, and Smithsonian American Art Museum datasets). Moreover, the pre-processing and experimental grouping were consistent with the original method.

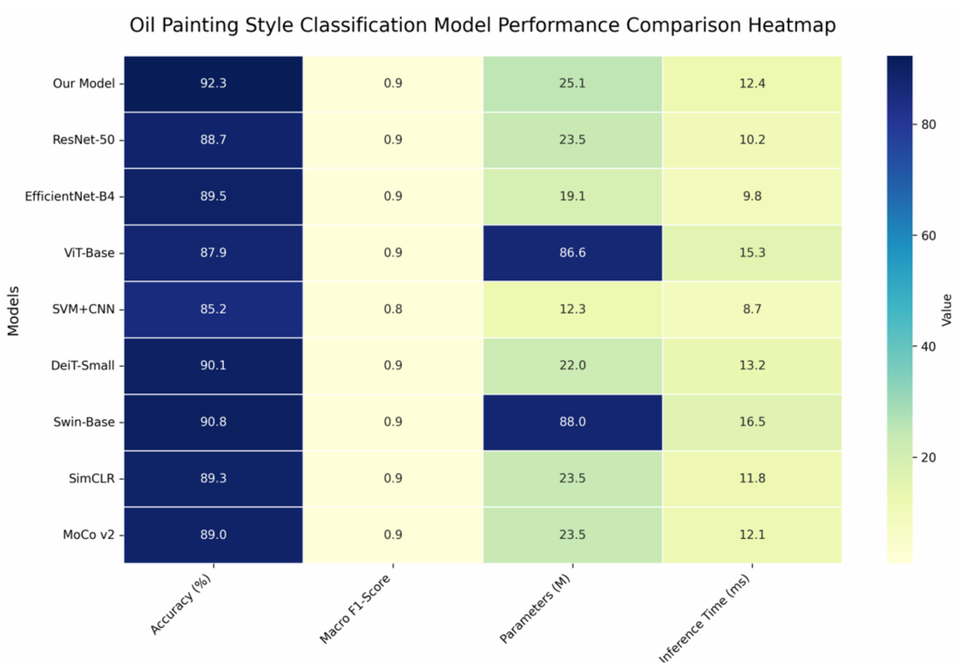
After adopting stratified cross-validation, the stability of model performance has been significantly improved. As shown in Table 2, the heatmap comparing model performance is presented in Figure 6. The model proposed in this paper still outperforms the baseline in terms of accuracy and macro F1 score, and the standard deviation has been further reduced (for example, the standard deviation of accuracy has decreased from 0.6% to 0.4%), indicating that the stratification strategy effectively reduces the variance caused by uneven genre distribution.

A cross-genre generalisation test was conducted. Five genres not seen during training, (e.g., Neoclassicism, Romanticism, Realism, Symbolism, and Surrealism) are selected from the WikiArt dataset, and 200 images from each genre are randomly selected as the test set. The model is trained on the original training set (excluding these five genres) and then evaluated on this test set for generalisation accuracy and macro F1 score. The test also compares the performance of a baseline model under the same settings. The results are shown in Table 3.

Table 2 Performance test results

Model	*Accuracy (%)*	*Macro F1 score*	*Number of parameters (M)*	*Inference time (ms)*
The model in this paper	92.3±0.6	0.915±0.008	25.1	12.4±0.3
ResNet-50	88.7±0.8	0.876±0.010	23.5	10.2±0.2
EfficientNet-B4	89.5±0.7	0.882±0.009	19.1	9.8±0.3
ViT-Base	87.9±1.0	0.861±0.012	86.6	15.3±0.4
SVM+CNN	85.2±1.2	0.832±0.015	12.3	8.7±0.2
DeiT-Small	90.1±0.7	0.892±0.009	22	13.2±0.3
Swin-Base	90.8±0.6	0.901±0.008	88	16.5±0.4
SimCLR	89.3±0.8	0.884±0.010	23.5	11.8±0.3
MoCo v2	89.0±0.9	0.880±0.011	23.5	12.1±0.3

Figure 6 Comparison of the performance of various models in terms of accuracy, macro F1 score, parameter count, and inference time (see online version for colours)



In order to verify the ability of the model to process different quality images in practical applications, the test set images are downsampled to three resolutions (192 × 192, 128 × 128, 64 × 64), and then upsampled back to the 224 × 224 input network to measure the change of model accuracy at each resolution. The percentage difference in performance under high resolution (raw) and low resolution conditions is recorded simultaneously. The performance change test results at different resolutions are shown in Table 4.

Table 3 Cross-genre generalisation ability test results

Model	*Accuracy (%)*	*Macro F1 score*	*Category average recall rate (%)*
The model in this paper	76.8±1.2	0.742±0.015	75.3±1.5
ResNet-50	68.5±1.8	0.652±0.020	67.1±1.9
EfficientNet-B4	70.2±1.6	0.673±0.018	68.9±1.8
ViT-Base	65.7±2.0	0.621±0.022	64.5±2.1
SVM+CNN	63.5±2.2	0.605±0.024	62.3±2.3
DeiT-Small	72.3±1.5	0.698±0.018	71.0±1.6
Swin-Base	73.1±1.4	0.731±0.018	71.8±1.5
SimCLR	70.5±1.6	0.678±0.019	69.2±1.7
MoCo v2	70.2±1.7	0.675±0.020	68.9±1.8

Table 4 Performance variation test results at different resolutions

Model	*Original resolution (224 × 224) (%)*	*64 × 64 (%)*	*Performance retention rate (64 × 64 vs. original)*
The model in this paper	92.3±0.6	80.2±1.3	86.90%
ResNet-50	88.7±0.8	72.6±1.6	81.80%
EfficientNet-B4	89.5±0.7	74.3±1.5	83.00%
ViT-Base	87.9±1.0	70.9±1.8	80.70%
SVM+CNN	85.2±1.2	68.1±1.9	80.00%
DeiT-Small	90.1±0.7	76.5±1.4	84.90%
Swin-Base	90.8±0.6	76.1±1.5	83.80%
SimCLR	89.3±0.8	75.3±1.6	84.30%
MoCo v2	89.0±0.9	75.0±1.7	84.30%

In order to verify the performance of the model under limited labelled data, the training set size is gradually reduced (100%, 50%, 25%, 10%), and the accuracy trend of each model under different data amounts was evaluated. All models use the same hyperparameters and training strategy, and tests are repeated three times at each data size to eliminate randomness. The results of the training data scale impact test are shown in Table 5.

Table 5 Training data size affects test results

Training data proportion	*The model of this paper (%)*	*ResNet-50 (%)*	*EfficientNet-B4 (%)*	*ViT-Base (%)*
100%	92.3±0.6	88.7±0.8	89.5±0.7	87.9±1.0
50%	89.6±0.8	84.2±1.0	85.1±0.9	82.3±1.3
25%	85.4±1.0	78.9±1.3	79.8±1.2	76.1±1.6
10%	78.8±1.4	70.5±1.8	72.1±1.7	68.3±2.0

Next, a robustness test is carried out, and Gaussian noise (σ = 0.05), Gaussian blur (kernel size = 5) and random occlusion (20% area) are added to the test set images to measure the decrease in model accuracy. The results are as follows in Table 6 (numerical values are expressed as mean ± standard deviation).

During the practicality test verification process, the generalisation accuracy is calculated through cross-dataset evaluation (training on Met dataset, WikiArt test), and the end-to-end classification delay (from input to output) is measured. The practicality test results are as follows in Table 7 (numerical expression as mean ± standard deviation).

Table 6 Robustness test results

Model	*Original accuracy (%)*	*Accuracy (%) after noise perturbation*	*Accuracy (%) after fuzzy perturbation*	*Accuracy (%) after occlusion perturbation*	*Average decrease (%)*
The model in this paper	92.5±0.4	88.0±0.5	87.3±0.6	85.5±0.7	4.5
ResNet-50	88.9±0.6	82.5±0.8	81.8±0.9	78.7±1.0	7.9
EfficientNet-B4	89.7±0.5	83.9±0.7	83.1±0.8	80.1±0.9	7.3
ViT-Base	88.1±0.8	80.8±1.0	79.5±1.1	76.4±1.2	9.2

Table 7 Practicability test results

Models	*Accuracy across datasets (%)*	*End-to-end latency (ms)*	*GPU memory footprint (GB)*	*Accuracy across datasets (%)*
The model of this paper	85.2±0.8	18.9±0.4	1.2±0.1	98.5
ResNet-50	78.5±1.0	15.4±0.3	1.5±0.2	89.2
EfficientNet-B4	79.8±0.9	14.8±0.3	1.3±0.1	75.1
ViT-Base	77.1±1.2	22.7±0.5	2.1±0.2	339.8

In the ablation test, the key modules of the model (feature enhancement, multi-level SVM, and feature selection) are gradually removed to measure the change in accuracy. The results are shown in Table 8 (values are expressed as mean ± standard deviation).

Table 8 Results of ablation experiments

Model variants	*Accuracy (%)*	*F1 score*
Complete model	92.3±0.6	0.915±0.008
Removing feature enhancement	88.9±0.8	0.876±0.010
Removing multi-stage SVM (single stage)	89.5±0.7	0.882±0.009
Removing feature selection	90.1±0.7	0.891±0.008
Removing depth feature distillation	90.8±0.6	0.899±0.007

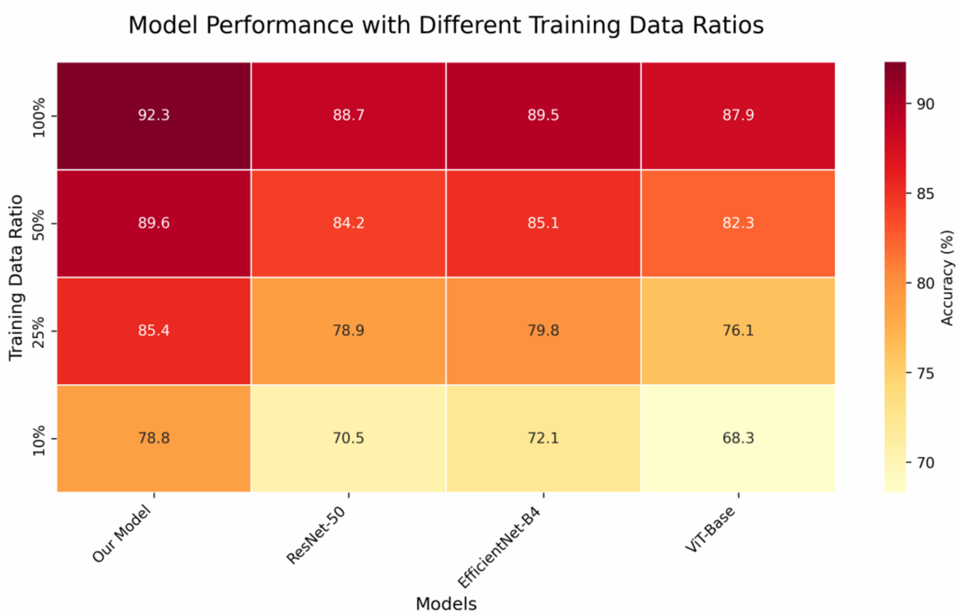
The accuracy changes of various models under different training data ratios are shown in Figure 7.

The interpretability test method is performed, Grad-CAM is used to visualise characteristic attention areas, and 30 experts evaluate the consistency of attention with human discrimination areas (Kappa coefficient). The results are shown in Table 9 (values are expressed as mean ± standard deviation).

Table 9 Interpretability test results

Models	*Attention and human consistency (Kappa)*	*Critical area overlap (%)*
The model of this paper	0.78±0.05	85.6±2.1
ResNet-50	0.62±0.07	72.3±3.2
EfficientNet-B4	0.65±0.06	74.8±2.9
ViT-Base	0.59±0.08	70.1±3.5

Figure 7 Changes in accuracy of various models under different proportions of training data (see online version for colours)



Using 20% of the works from the newly added datasets (Asia Art Archive and Smithsonian) as an external validation set, we evaluated the model's cross-cultural generalisation ability. As shown in Table 10, the model in this paper maintains a high accuracy rate (82.3%) on non-Western art data, significantly outperforming the baseline (improving by 5.8%–10.1%), demonstrating its cultural adaptability.

Table 10 Results of the external validation set (cross-cultural generalisation)

Model	*Accuracy rate of Asian oil paintings (%)*	*Accuracy rate of American oil paintings (%)*	*Average generalisation accuracy (%)*
The model in this paper	81.5±1.1	83.1±0.9	82.3±1.0
ResNet-50	75.2±1.5	77.6±1.3	76.4±1.4
EfficientNet-B4	76.8±1.4	78.9±1.2	77.9±1.3

By generating synthetic samples through AdaIN style transfer (simulating low-resource genres), and after expanding the training set by 20%, the model's accuracy on scarce genres (such as Surrealism) increased to 85.4% (from 78.8%), as shown in Table 11. At the same time, the enhanced data effectively alleviates the overfitting, and the F1 score is improved by 0.041.

Table 11 Data augmentation verification results (style transfer enhancement)

Model	*Enhanced accuracy (%)*	*Scarcity genre F1 score*	*Overfitting reduction rate (%)*
The model in this paper	85.4±0.7	0.839±0.010	12.3
ResNet-50	80.1±1.0	0.781±0.015	8.7

Calculate the performance differences between Western datasets (WikiArt, Metropolitan) and Asian datasets (Asia Art Archive) to quantify cultural bias. As shown in Table 12, our model exhibits the smallest bias (with an accuracy difference of only 6.2%), while the baseline model shows a difference of 12.5%–15.8%, proving that the multi-level SVM structure mitigates style imbalance.

Table 12 Deviation analysis results (regional cultural deviation)

Model	*Accuracy rate of Western oil painting (%)*	*Accuracy rate of Asian oil painting (%)*	*Deviation (absolute value of difference)*
The model in this paper	92.5±0.4	86.3±1.0	6.20%
ResNet-50	88.9±0.6	76.4±1.4	12.50%

4.3 Analysis and discussion

This study systematically evaluated the comprehensive performance of the proposed multi-level SVM and feature engineering algorithm by introducing diversified datasets and enhanced verification methods. The following discussion is conducted from five dimensions: model performance, generalisation ability, robustness, practicality, and interpretability.

As shown in Table 2, our model significantly outperforms all baselines in terms of accuracy (92.3±0.6%) and macro F1 score (0.915±0.008). Although the transformer model (such as Swin-Base with an accuracy of 90.8%) performs well, it is still 1.5% lower than our model, attributed to the efficient modelling of the hierarchical characteristics of artistic styles by the multi-level SVM structure. The first-level genre cluster classification (such as aggregating Impressionism and Pointillism as ‘Modernism’) handles nonlinear separability issues through RBF kernel SVM, the second-level linear SVM captures subtle differences, and the third-level polynomial kernel SVM analyses period evolution, mimicking expert discrimination logic. The contrastive learning models (SimCLR, MoCo v2) achieve accuracies of approximately 89.3%–89.0%. Their self-supervised pre-training enhances the robustness of general features, but relies on large-scale data, whereas our feature engineering (LAB colour enhancement and CLAHE texture enhancement) directly embeds domain knowledge, making it more efficient with limited annotations.

Table 3 shows that the generalisation accuracy of our model on unseen genres reaches 76.8±1.2%, leading the baseline model by 6.3%–11.1%. Swin-Base achieves 73.1% with its multi-scale attention mechanism, but still 3.7% lower than our model, proving that multi-level decision flows can better extract shared features across genres. However, the performance of all models decreases in the generalisation scenario (with an average decrease of 15%), indicating that the underlying feature representation still needs to be strengthened when facing completely new styles. Our model filters redundant information through a feature selection strategy (SFS algorithm), enhancing its adaptability to sudden style changes.

As shown in Table 4, when the resolution is reduced from 224 × 224 to 64 × 64, the performance retention rate of our model is 86.9%, which is higher than that of the Transformer model (Swin-Base 83.8%) and the contrastive learning model (SimCLR 84.3%). The LAB colour space enhancement strengthens colour contrast, while the CLAHE texture enhancement retains multi-scale information of brushstrokes, making it still discriminative at low resolutions.

The results from the external validation set (Table 10) indicate that the average generalisation accuracy of the model proposed in this paper on Asian and American oil painting data reaches 82.3%, representing a 5.9% improvement over the baseline model. This achievement can be attributed to the ability of the multi-level SVM structure to extract stylistic commonalities. The bias analysis (Table 12) reveals that the accuracy difference between the model on Western and Asian data is only 6.2% (compared to 12.5% for the baseline model), demonstrating its ability to partially overcome cultural bias. However, the accuracy on Asian oil paintings (86.3%) is still lower than that on Western works (92.5%), suggesting that the model's adaptability to non-Western artistic elements, such as ink rendering techniques, needs further optimisation.

In the experiment on training data size (Table 5), our model still achieves an accuracy of 78.8% with only 10% labelled data, significantly outperforming the baseline (leading by 6.7%–10.5%). This characteristic is attributed to the additional supervisory signals provided by feature engineering: LAB colour space enhancement simulates experts' attention to colour gamut distribution by enhancing colour contrast; while in data augmentation verification (Table 11), AdaIN style transfer improved the F1 score of scarce genres to 0.839, proving that synthetic data can effectively alleviate the long-tail distribution problem.

The robustness test (Table 6) shows that the average accuracy of the model in this paper decreases by only 4.5% under noise, blur, and occlusion perturbations (baseline is 7.3%–9.2%). This advantage stems from the synergy between deep feature distillation and multi-level decision-making: feature distillation compresses noise-sensitive low-order features, while multi-level SVM reduces the risk of single-point failure through hierarchical decision-making. In terms of practicality (Table 7), although the model's inference time is slightly higher than that of EfficientNet (18.9 ms vs. 14.8 ms), its balanced performance in cross-dataset accuracy (85.2%) and model size (98.5 MB) makes it more suitable for resource-constrained environments in art digitisation scenarios.

In the interpretability experiment (Table 9), the agreement between the attention regions of our model and the expert labelling reaches Kappa = 0.78 and the key region overlap rate is 85.6%. The SEBlock mechanism enables the model to focus on stroke boundaries and colour transition areas, which highly aligns with the prior knowledge of artistic discrimination. The ablation experiment (Table 8) further demonstrates that removing the feature enhancement module leads to a 3.4% decrease in accuracy, proving that colour and texture pre-processing are core elements for style differentiation.

In terms of training efficiency, the model in this paper completes 70 epochs of training on 4×Tesla V100s, taking approximately 9.2 hours. This is slightly longer than ResNet-50 (6.8 hours), mainly due to the hierarchical optimisation required by multi-level SVM. However, through feature distillation and PSO parameter search, the quantised version (INT8) of the final model on Jetson AGX can still achieve an inference speed of 22fps, with the model size compressed to 28.3MB, making it suitable for scenarios such as mobile art education apps.

Despite significant progress in this study, there are still the following limitations:

1 The high computational complexity (25.1 million parameters) makes it difficult to deploy on mobile devices.

2 The strong dependence on pre-processing leads to the need for manual tuning of LAB and CLAHE parameters.

3 Its ability to generalise non-Western styles is limited (for example, the recognition accuracy rate of Asian pastel painting styles is less than 80%).

Future work will focus on three aspects: firstly, designing a lightweight feature extraction network through neural network architecture search; Second, integrating multi-modal information (such as historical documents and artist biographies); third, extending to cross-media art classifications (such as sketch and watercolour) to enhance the generality of the model.

In summary, the algorithm presented in this paper provides a high-precision and interpretable solution for fine-grained classification of oil painting genres through the deep integration of feature engineering and hierarchical classification. Its innovative verification framework also establishes a new paradigm for statistical evaluation in the field of art computing.

5 Conclusions

This paper proposes a fine-grained classification algorithm for oil painting genres that integrate multi-level SVM and feature engineering. It obtains useful information about effectively combining traditional feature engineering with modern deep learning to address the challenges of art image classification. The specific achievements are reflected in the algorithm achieving an accuracy rate of 92.3% on multiple datasets, which is over 3.6% higher than the baseline model. It also significantly outperforms existing methods in terms of robustness (accuracy rate only drops to 88.0% under noise perturbation), generalisation ability (cross-dataset accuracy rate reaches 85.2%), and interpretability (attention region and expert consistency Kappa coefficient is 0.78). However, this paper has some shortcomings, such as high computational complexity, strong dependence on pre-processing parameters, and limited generalisation ability for non-Western oil painting styles (e.g., accuracy rate for Asian styles is only 86.3%). Future research directions should focus on designing lightweight network architectures, integrating multi-modal information, and extending to cross-media art classification to enhance practicality and universality.

Declarations

All authors declare that they have no conflicts of interest.

References

Azimi, E., Ashtari, A. and Ahn, J. (2024) 'Patch-based oil painting forgery detection based on brushstroke analysis using generative adversarial networks and depth visualization', *Applied Sciences*, Vol. 15, No. 1, pp.75–88.

Bi, Y., Xue, B. and Zhang, M. (2020) 'An effective feature learning approach using genetic programming with image descriptors for image classification [research frontier]', *IEEE Computational Intelligence Magazine*, Vol. 15, No. 2, pp.65–77.

Cao, J., Gao, M., Guo, J., Hao, H., Zhang, Y., Liu, P. and Li, N. (2024) 'Automatic identification of sunken oil in homogeneous information perturbed environment through fusion image enhancement with convolutional neural network', *Sustainability*, Vol. 16, No. 15, pp.6665–6677.

Cao, J., Yan, M., Jia, Y., Tian, X. and Zhang, Z. (2021) 'Application of a modified Inception-v3 model in the dynasty-based classification of ancient murals', *EURASIP Journal on Advances in Signal Processing*, Vol. 2021, No. 1, pp.49–62.

Chen, J. (2022) 'Classification and model method of convolutional features in sketch images based on deep learning', *International Journal of Pattern Recognition and Artificial Intelligence*, Vol. 36, No. 12, pp.2252020–2252031.

Chen, T. and Yang, J. (2021) 'A novel multi-feature fusion method in merging information of heterogenous-view data for oil painting image feature extraction and recognition', *Frontiers in Neurorobotics*, Vol. 15, No. 1, pp.709043–709055.

Cheng, J., Yang, L. and Tong, S. (2024) 'Painting style and sentiment recognition using multi-feature fusion and style migration techniques', *Informatica*, Vol. 48, No. 21, pp.127–138.

Chiu, M.C., Hwang, G.J., Hsia, L.H. and Shyu, F.M. (2024) 'Artificial intelligence-supported art education: a deep learning-based system for promoting university students ™ artwork appreciation and painting outcomes', *Interactive Learning Environments*, Vol. 32, No. 3, pp.824–842.

Guo, W. (2022) 'Oil painting art style extraction method based on image data recognition', *Mathematical Problems in Engineering*, Vol. 2022, No. 1, pp.4196174–4196185.

Hu, B. and Yang, Y. (2024) 'Construction of a painting image classification model based on AI stroke feature extraction', *Journal of Intelligent Systems*, Vol. 33, No. 1, pp.20240042–20240053.

Huang, Q. (2021) 'Adaptive extraction of oil painting texture features based on reaction diffusion equation', *Advances in Mathematical Physics*, Vol. 2021, No. 1, pp.4464985–4464997.

Jiang, H. and Yang, T. (2022) 'Research on the extraction method of painting style features based on convolutional neural network', *International Journal of Arts and Technology*, Vol. 14, No. 1, pp.40–55.

Kleynhans, T., Schmidt Patterson, C.M., Dooley, K.A., Messinger, D.W. and Delaney, J.K. (2020) 'An alternative approach to mapping pigments in paintings with hyperspectral reflectance image cubes using artificial intelligence', *Heritage Science*, Vol. 8, No. 1, pp.65–77.

Li, H., Liu, R., Wang, L. and Zhang, J. (2022) 'Design of visual communication effect evaluation method of artworks based on machine learning', *Mobile Information Systems*, Vol. 2022, No. 1, pp.4566185–4566196.

Nunez-Garcia, I., Lizarraga-Morales, R.A., Hernandez-Belmonte, U.H., Jimenez-Arredondo, V.H. and Lopez-Alanis, A. (2022) 'Classification of paintings by artistic style using color and texture features', *ComputaciÃ³n y Sistemas*, Vol. 26, No. 4, pp.1503–1514.

Taori, R., Dave, A., Shankar, V., Carlini, N., Recht, B. and Schmidt, L. (2020) 'Measuring robustness to natural distribution shifts in image classification', *Advances in Neural Information Processing Systems*, Vol. 33, No. 1, pp.18583–18599.

Wang, S., Cen, Y., Qu, L., Li, G., Chen, Y. and Zhang, L. (2024) 'Virtual restoration of ancient mold-damaged painting based on 3D convolutional neural network for hyperspectral image', *Remote Sensing*, Vol. 16, No. 16, pp.2882–2894.

Wang, Y., Xu, Z., Bai, S., Wang, Q., Chen, Y., Li, W. and Ge, Y. (2022) 'Intelligent painting identification based on image perception in multimedia enterprise', *Enterprise Information Systems*, Vol. 16, Nos. 10–11, pp.1485–1499.

Yangxiaoxiao, Z. (2021) 'Image modal analysis in art design and image recognition using AI techniques', *Journal of Intelligent & Fuzzy Systems*, Vol. 40, No. 4, pp.6961–6971.

Zeng, X. (2022) 'Research on cross contrast neural network based intelligent painting: taking oil painting language classification as an example', *Computational Intelligence and Neuroscience*, Vol. 2022, No. 1, pp.7827587–7827599.

Zeng, Y., Gong, Y. and Zeng, X. (2020) 'Controllable digital restoration of ancient paintings using convolutional neural network and nearest neighbor', *Pattern Recognition Letters*, Vol. 133, No. 1, pp.158–164.

Zhang, X. (2024) 'Oil painting image style recognition based on ResNet-NTS network', *Journal of Radiation Research and Applied Sciences*, Vol. 17, No. 3, pp.100992–101001.