applied sciences

MDPI

Article

Understanding Regional and Stylistic Diversity in Chinese Rural Paper-Cutting Through Convolutional Neural Network-Based Image Classification

Xiaochu Wu [1], Xiaoyue Yin [2], Xiaofeng Chen [2], Xudong You [2], Fang Zhang [1,*] and Yi Xiao [3,*]

[1] HNU-ASU International College, Hainan University, Haikou 571155, China; 20253004948@hainanu.edu.cn
[2] State Key Laboratory of Chemo and Biosensing, College of Chemistry and Chemical Engineering, Hunan University, Changsha 410082, China; yinxy@hnu.edu.cn (X.Y.); chenxf05@hnu.edu.cn (X.C.); xvdongyou@hnu.edu.cn (X.Y.)
[3] College of Fine Arts, Hunan Normal University, Changsha 410081, China
* Correspondence: fzhang_haitc@hainanu.edu.cn (F.Z.); xiaoyi007@hunnu.edu.cn (Y.X.)

Abstract

As an important component of Chinese folk art, rural paper-cutting embodies rich regional cultural connotations and distinctive aesthetic expressions. In this study, a Chinese rural paper-cutting image dataset covering multiple regions and artistic styles was constructed, and a convolutional neural network (CNN)-based framework was proposed for regional and stylistic identification of paper-cutting works. Five representative mainstream CNN models were evaluated for both tasks. For regional classification, all models achieved high accuracy, with EfficientNet-B1 attaining the highest accuracy of 91.46%. The style classification task was more challenging due to subtle visual differences, with MobileNetV3-Small achieving the highest accuracy of 73.20%. In addition, t-distributed stochastic neighbor embedding (t-SNE) visualizations further confirmed that the models were able to effectively distinguish different regional and stylistic categories in high-dimensional space. To enhance model interpretability, Gradient-weighted Class Activation Mapping (Grad-CAM) was applied to visualize the optimal models. The results show that the CNNs consistently focus on core structural features of paper-cutting works, suggesting that CNNs can capture visually and culturally meaningful features. Overall, this study demonstrates the feasibility of applying CNNs to the analysis of traditional folk art and provides a practical technical pathway for digital management, intelligent classification, and educational dissemination of rural paper-cutting art.

Keywords: Chinese paper-cutting; convolutional neural networks; computer vision; style classification; regional art styles

Check for updates

Received: 1 March 2026
Revised: 20 March 2026
Accepted: 23 March 2026
Published: 25 March 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) license.

1. Introduction

Paper-cutting, also known as paper carving, first appeared in Chinese folk celebrations [1]. As the most widely disseminated form of Chinese folk art, paper-cutting is a visual cultural practice created and transmitted by the general public through long-term production and everyday life. Chinese paper-cutting is essentially a two-dimensional decorative art form, in which patterns are created by cutting or carving flat sheets of paper. It is characterized by strong regional specificity, a broad popular foundation, and distinctive aesthetic features [2]. Through techniques such as cutting, carving, and chiseling, paper is transformed into symbolically meaningful visual signs and compositional carriers of

artistic form. These works are extensively used in festival decorations, blessing rituals, and daily life, embodying people's understanding and artistic expression of natural order, social life, and spiritual beliefs [3]. Folk paper-cutting artists employ this medium as a vehicle for emotion and lived experience, distilling aspirations for favorable weather, abundant harvests, and a good life into a visual artistic language. In this way, paper-cutting has become an important means of preserving rural cultural memory and expressing idealized desires [4].

Influenced by diverse geographical environments, historical conditions, and cultural traditions, Chinese paper-cutting has gradually developed a wide range of artistic styles with distinct regional cultural characteristics and popular aesthetic preferences [5]. Over time, prominent geographical landscapes have been continuously integrated into everyday visual experience, shaping local paper-cutting styles. In northern regions, the rugged power of wind and sand in frontier landscapes has been interpreted and transformed into bold, forceful, and dynamic paper-cutting forms. In contrast, the southern regions, characterized by rich vegetation and year-round lushness, have fostered an aesthetic appreciation for intricate details of growth and vitality, which is reflected in paper-cutting works featuring intricate knife techniques, delicate layered structures, and continuous flowing lines imbued with a sense of life force. The harmonious relationship between humans and nature, rooted in the concept of "unity of heaven and humanity," provides inspiration for artistic creation, resulting in regional differences in composition, line organization, and modeling tendencies [6]. Paper-cutting works from northern areas (e.g., Hebei, Shanxi, and Shaanxi) are typically valued for their simplicity, strength, and rugged expressiveness, with powerful forms and strong dynamic tension. By contrast, works from southern regions (e.g., Guangdong, Jiangsu, and Zhejiang) tend to be slender, refined, and visually elaborate, exhibiting delicate and sophisticated characteristics [7]. Such stylistic differences not only reflect the diversity of popular aesthetic preferences but also reveal the deep influence of human–land relationships on artistic expression. Based on artistic techniques and expressive features, paper-cutting can be broadly classified into rugged, intricate, concise, and delicate styles [3]. Among these, the rugged style prevalent in northern paper-cutting is particularly representative, characterized by thick lines, simplified forms, an emphasis on expressive resemblance rather than detailed realism, and a free, unrestrained approach [8]. In southern regions, the intricate style is distinguished by the use of openwork cutting techniques to create graceful, finely detailed lines and elaborate patterns, reflecting a strong preference for refinement and precision [6]. The concise style, commonly found in northern paper-cutting, emphasizes clear and fluent outlines of main forms, with minimal decorative details, highlighting the symbolic and functional role of paper-cutting. Such works are often used in folk practices such as blessings, vow fulfillment, and ritual ceremonies to convey strong emotional intentions and religious beliefs [5]. The delicate style is more frequently observed in southern paper-cutting works, featuring highly detailed designs, rich layering, and auspicious motifs. Through dense and refined ornamentation, these works create a solemn and festive visual atmosphere and are commonly used for celebrations and decorative purposes [2]. These classifications are of significant value in art historical and folkloristic studies, as they help reveal how paper-cutting responds to specific cultural contexts and lived experiences through symbolic imagery and visual metaphors [9].

Chinese paper-cutting has been recognized as an important component of the world's intangible cultural heritage [10]. However, it currently faces a dual challenge: a generational gap among practitioners and a decline in cultural influence in the information age. Once widely practiced and transmitted, paper-cutting now encounters stagnation in creative development and a contraction of public cultural impact. As the scholar Feng Jicai has warned, "It is rapidly disappearing and disintegrating; we must protect it." The survival

and development of folk culture depends fundamentally on inheritors and the effectiveness of protection and dissemination mechanisms [11]. Introducing paper-cutting into children's aesthetic education enables young learners to engage with multi-ethnic cultures through hands-on practice, while naturally sustaining and reinterpreting regional paper-cutting forms and symbolic systems. In this sense, rural paper-cutting serves not only as a key medium for intergenerational transmission of folk art but also as a concrete practice through which children participate in the construction of local culture. From the perspective of aesthetic education, paper-cutting holds irreplaceable value. As an art form with low material and technical entry barriers but high learner engagement, paper-cutting plays a positive role in cultivating aesthetic perception, manual skills, and cultural identity among children [12,13]. Therefore, systematic analysis of rural paper-cutting art can contribute to a deeper understanding of folk art transmission pathways, regional cultural influences, and directions in children's aesthetic education.

Art digitization can promote the transmission of cultural heritage [14]. However, at the data level, current research lacks systematically organized and well-annotated image resources that explicitly distinguish regional and stylistic attributes in Chinese rural paper-cutting. Existing research mainly focuses on paintings or general art images, lacking resources for analyzing the unique line-cutting visual structure of paper-cutting. Furthermore, due to the reliance on individualized artistic sensibility and experiential judgment, traditional studies of paper-cutting styles have long depended on expert knowledge. As a result, style identification and typological classification are inevitably influenced by researchers' subjective interpretations, leading to certain limitations [15]. In recent years, advances in computer vision and deep learning have prompted the gradual introduction of image-based analytical methods into art and humanities research [16]. Convolutional neural networks (CNNs) are among the most representative models for image analysis in deep learning, with a core advantage in their ability to automatically learn local and global visual features through multi-layer convolutional and nonlinear mapping structures [17]. Unlike traditional approaches that rely on manually designed features, CNNs can hierarchically extract edges, textures, morphological structures, and compositional relationships directly from raw image data, forming multi-level representations of image content [18]. This capability makes CNNs particularly well suited to handling artistic images with complex visual structures and subtle stylistic differences. Although previous studies have applied CNN models to painting style recognition and folk art analysis, demonstrating good classification performance [19–21], few studies have focused on the classification of folk art with cultural semantics. In particular, the classification of geographical origin and artistic style has not been fully explored. Furthermore, existing research often focuses on classification performance while lacking interpretability analysis of the model's decision-making process, which is detrimental to the understanding and inheritance of artworks.

Based on the above issues, in this study, Chinese rural paper-cutting artworks were taken as the research object, and CNN-based algorithms were employed to automatically classify their regional and stylistic characteristics, complemented by explainable methods to analyze the model decision-making process. The overall experimental workflow is illustrated in Figure 1. First, a curated collection of paper-cutting images covering multiple regions and styles was compiled to support the computational analysis in this study. Second, five representative CNN models were selected to perform systematic regional and style classification, along with comparative performance evaluation, in order to assess their suitability for paper-cutting style classification tasks. Finally, explainable analysis of the classification results was conducted to reveal, at the feature level, the morphological structures and pattern elements relied upon by deep learning models during paper-cutting image discrimination. This approach offers a new perspective for understanding artistic

style differences in Chinese rural paper-cutting. The proposed artificial intelligence models are expected to be further integrated into digital exhibition platforms, educational support systems, and cultural heritage management systems, enabling automatic annotation, classification-based retrieval, and intelligent recommendation of paper-cutting works. At the same time, they support the living transmission of folk life aesthetics and localized inheritance of traditional Chinese culture through children's aesthetic education, providing an interpretable intelligent analytical basis for the visual understanding of regional cultural imagery.

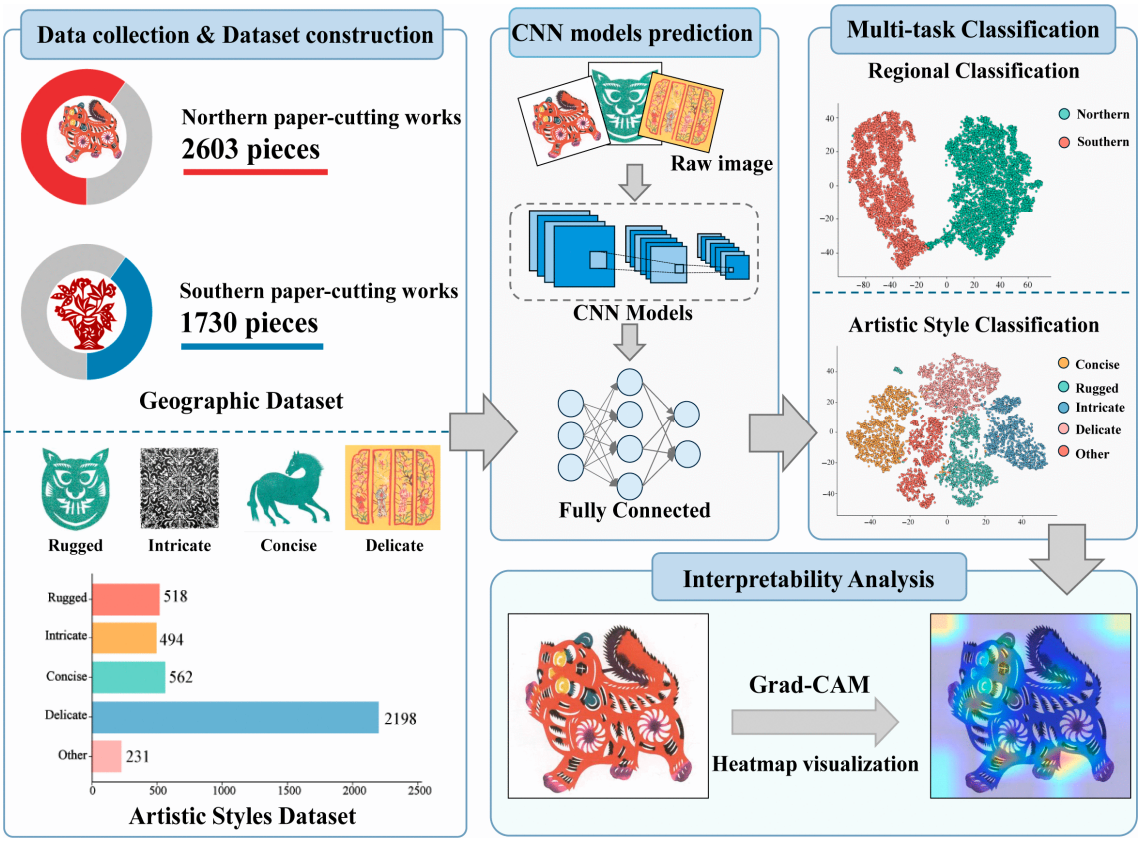


Figure 1. Workflow of this work.

2. Materials and Methods

2.1. Dataset Construction

The effectiveness of machine learning-based approaches largely depends on the quality and representativeness of the training data. Without a well-constructed dataset, even advanced algorithms may fail to produce reliable or meaningful results [22]. Compared with general image classification tasks, publicly available image resources related to Chinese rural paper-cutting remain limited, particularly those with explicit regional and stylistic annotations. To facilitate the experiments in this study, a collection of paper-cutting images was assembled and organized from multiple sources, covering a range of regions and stylistic categories.

Specifically, a total of 4333 paper-cutting images from northern and southern regions of China were collected and organized for the regional classification task. This paper adopts a Southern/Northern dichotomy framework not to simplify the complex regional differences within China, but to examine the differentiation of artistic styles within the macro-human–land relationship. This division has a clear theoretical basis. On the one hand, the Qinling–Huaihe line has long been regarded as an important marker of the North–South divide in

geography, with the Qinling-Daba Mountains forming the main transitional zone between North and South. On the other hand, in the Chinese intangible cultural heritage system, folk paper-cutting is often summarized into two major types: Northern style and Southern style, corresponding to the overall characteristics of bold and unrestrained versus delicate and elegant, respectively. Furthermore, the long-standing aesthetic categories in Chinese art history, such as "Southern softness and Northern strength" and "Southern elegance and Northern grandeur," provide stable theoretical support for the adoption of this framework in this study. During data collection, each image was matched as closely as possible to its actual place of origin (such as province or specific region), and combined with the common classifications in Chinese natural geography and folk art research, the samples were categorized into Southern or Northern categories.

In addition, 4003 images were collected, curated, and annotated according to artistic style. These images were categorized into five style classes, including rugged, intricate, concise, delicate, and other styles. In this study, the classification of paper-cutting styles is grounded in established frameworks from Chinese folk art research, particularly drawing on the typological perspectives proposed by Feng Jicai in his work on the classification of Chinese folk art [23]. Based on these theoretical foundations and expert consultation, four primary styles were defined as the main categories: rugged, intricate, concise, and delicate.

However, during dataset annotation, it was observed that a subset of paper-cutting works could not be reliably assigned to any single predefined category. This is mainly due to two reasons. First, with increasing cultural exchange and artistic development, many contemporary paper-cutting works exhibit hybrid characteristics, combining elements from multiple traditional styles. Second, some works demonstrate strong individual creativity or experimental features that do not conform to conventional stylistic definitions. To ensure labeling consistency and avoid forcing ambiguous samples into inappropriate categories, these works were grouped into the "Other" category. This category therefore represents stylistically mixed, transitional, or non-standard works that fall outside the four primary classes.

To minimize the impact of individual experience differences, this paper introduces a multi-expert collaboration mechanism in the data annotation stage. Specifically, three practitioners with over ten years of experience in paper-cutting creation and teaching were invited to participate in the annotation work. Each image was independently judged by the three experts to best fit its style type. When the three experts agreed, the label was directly assigned; when there was disagreement, it was retained after reaching a consensus through group discussion; and if no consensus could be reached, the sample was removed from the dataset. This process ensures the consistency and reliability of the labels as much as possible, while reducing the interference of ambiguous samples on model training.

Paper-cutting images from different regions and styles together constitute a relatively comprehensive dataset of Chinese rural paper-cutting artworks, designed to support a systematic analysis of regional and stylistic diversity in Chinese rural paper-cutting art. The distribution of samples across different regions and style categories is illustrated in Figure 2. For both the regional and style datasets, the images were split into training, validation, and test sets using an 8:1:1 ratio.

2.2. Image Preprocessing

To ensure compatibility between the input images and the CNN architectures, a standardized image preprocessing pipeline was adopted in this study. The primary objective of preprocessing was to normalize the input image format while preserving the visual features essential for artistic analysis. First, all input images were resized to a fixed spatial resolution of 224 × 224 pixels. This resolution is widely used in CNN-based image classification

tasks and is compatible with the input requirements of most modern CNN architectures. Standardizing image size facilitates batch processing and stable model training, while retaining sufficient visual information related to the contours, textures, and structural details of paper-cutting artworks. Moreover, this resolution provides a reasonable balance between representational fidelity and computational efficiency.

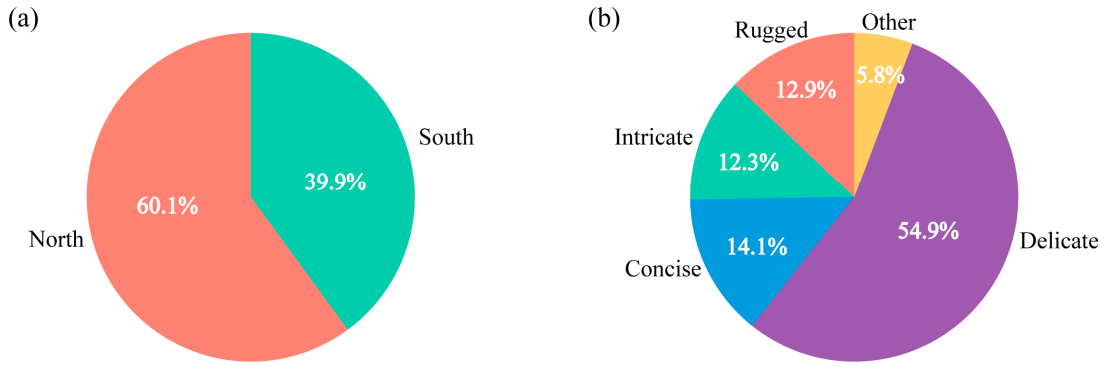


Figure 2. Distribution of the Chinese rural paper-cutting image dataset. (**a**) Distribution of samples across different regions; (**b**) distribution of samples across different artistic styles.

To improve model generalization, reduce the risk of overfitting, and mitigate class imbalance, data augmentation was applied exclusively to the training set. No augmentation was performed on the validation or test sets to ensure a fair evaluation of model performance. To partially address class imbalance, augmentation was selectively applied to underrepresented classes, resulting in a more balanced training distribution. Considering the symbolic nature and geometric symmetry of paper-cutting artworks, only geometric transformations were employed. Specifically, two types of random geometric transformations were applied:

- Random rotation. Given the variation in line orientation across paper-cutting works, each training image was randomly rotated within a restricted range, with the rotation angle selected as an integer multiple of 90°:

$$I' = R(\theta)\, I, \tag{1}$$

where I represents the original image and I' the transformed image, and $R(\theta)$ is the rotation matrix, $\theta \in \{0°, 90°, 180°, 270°\}$. This operation preserves the intrinsic structural characteristics of paper-cutting patterns and helps the model learn orientation-invariant features.

- Random flipping. As paper-cutting artworks often exhibit symmetrical compositions, each training image was randomly flipped with a probability of $p = 0.5$. Horizontal flipping (F_h) reflects the image about the vertical axis, while vertical flipping (F_V) reflects it about the horizontal axis:

$$I' = F_h\,(I), \tag{2}$$

$$I' = F_V\,(I), \tag{3}$$

where I represents the original image and I' represents the transformed image. This operation enables the model to better capture symmetry, a common characteristic of folk paper-cutting designs, and further enhances robustness to variations in orientation.

2.3. CNN Models

Before employing deep learning methods, we attempted to classify Chinese folk paper-cutting using traditional machine learning methods (SVM). However, the results showed that the classification performance of traditional methods was relatively poor, especially in more complex style classification tasks (Table S3). This indicates that traditional machine learning methods cannot capture the complex structural and stylistic features of paper-cutting artworks. Deep learning is now widely used in the art field, and existing research has shown that its performance far surpasses that of traditional machine learning methods [24]. Therefore, this study focuses on applying mature CNN models.

In this study, five representative CNN architectures were selected as visual analysis tools for classifying Chinese rural paper-cutting images. All selected models have been widely adopted in the computer vision field and have demonstrated robust performance across image recognition tasks in various application domains [25,26]. Rather than proposing new network architectures, this study focuses on applying well-established CNN models to investigate their effectiveness in capturing the visual characteristics embedded in traditional folk art images. A brief description of each model used in this study is provided below.

2.3.1. EfficientNet

EfficientNet, introduced in 2019, is a convolutional neural network based on the concept of compound scaling [27]. This approach was designed to address the long-standing trade-off between model size, accuracy, and computational efficiency. Unlike conventional methods that scale network depth, width, or input resolution independently, EfficientNet uniformly scales all three dimensions using a set of fixed scaling coefficients. This strategy enables the model to achieve improved accuracy while maintaining parameter efficiency, making it a preferred architecture for a wide range of challenging tasks, including object recognition, image segmentation, and even language processing [26,28,29]. In this study, two variants from the EfficientNet family, EfficientNet-B1 and EfficientNet-B4, were employed for modeling.

2.3.2. MobileNet

MobileNet, proposed in 2017, is a class of lightweight CNN architectures specifically designed for mobile and embedded vision applications [30]. Its core design is based on depthwise separable convolutions, which substantially reduce computational cost and parameter count while preserving essential feature extraction capabilities. This study adopted MobileNetV3-Small, an enhanced version that integrates neural architecture search and attention mechanisms to strengthen feature representations [31]. Owing to its compact structure and high computational efficiency, MobileNetV3-Small has become a representative lightweight model for image processing tasks [32,33].

2.3.3. MobileViT

MobileViT, introduced in 2020, is a hybrid architecture that combines CNNs and Transformers [34]. It leverages CNNs to extract local features while employing Vision Transformers (ViTs) to capture global representations. Under a limited parameter budget, MobileViT achieves higher accuracy than MobileNets and other CNN models while maintaining low computational complexity [35]. In this study, two variants, MobileViT-XX-Small and MobileViT-X-Small, were selected.

All models were trained under a unified experimental framework to ensure fair comparison. Specifically, the CNN models were implemented using transfer learning with ImageNet pre-trained weights, followed by fine-tuning on the paper-cutting dataset.

Details regarding the optimizer, learning rate tuning, number of training epochs, and early stopping criteria are provided in the Supplementary Information. In short, all models are trained using the Adam optimizer with an initial learning rate of 0.001. The early stopping strategy is based on validation loss, and the patience value is 10 cycles in the North South classification and 50 cycles in the style classification. Each experiment was repeated five times using different random seeds (42, 123, 456, 789, 2024) to eliminate the influence of model initialization differences, and average performance was reported. The CNN models were used for two types of image classification tasks: regional classification (northern vs. southern regions) and style classification of paper-cutting artworks. Detailed hyperparameter settings are provided in the Supplementary Information.

2.4. Evaluation Metrics

To evaluate the predictive performance of the model, we selected several effective metrics: accuracy (ACC), precision (Pre) and the area under the receiver operating characteristic curve (ROC-AUC). The equations for the first two metrics are expressed as follows:

$$\mathrm{ACC} = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{TN} + \mathrm{FP} + \mathrm{FN}} \tag{4}$$

$$\mathrm{Pre} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}} \tag{5}$$

where TP (true positive), TN (true negative), FP (false positive), and FN (false negative) represent the corresponding counts for each class. The values of these metrics range from 0 to 1, and higher values indicate better model performance.

In addition, we used the area under the receiver operating characteristic curve (ROC-AUC) to evaluate model performance. The ROC-AUC score is a commonly used metric for binary classifiers and represents the area under the ROC curve. The ROC curve plots the true positive rate (TPR) against the false positive rate (FPR), illustrating the classifier's performance across different decision thresholds. The ROC-AUC score ranges from 0 to 1, where an ROC-AUC of 0.5 indicates poor performance (equivalent to random guessing), and an ROC-AUC of 1 signifies excellent performance with perfect classification. The formulas for calculating TPR and FPR are as follows:

$$\mathrm{TPR} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \tag{6}$$

$$\mathrm{FPR} = \frac{\mathrm{FP}}{\mathrm{FP} + \mathrm{TN}} \tag{7}$$

3. Results

3.1. Regional Classification in Paper-Cutting Images

To evaluate the effectiveness of CNN in distinguishing regional characteristics in paper-cutting artworks, five CNN models were applied to the regional classification task. For each model, we repeated the experiments using five different random seeds, and the results are listed in Table 1 as mean and standard deviation. The best classification performance for each model on test set is shown in Figure 3a. Among all evaluated architectures, EfficientNet-B1 achieved the highest performance, with an accuracy of 91.46%. The remaining models also performed well, with accuracies of 90.54% for EfficientNet-B4, 90.24% for MobileNetV3-Small, and 89.33% for both MobileViT-XX-Small and MobileViT-X-Small. These results indicate that all selected CNN models are capable of achieving high classification performance on the constructed dataset, suggesting that region-related visual differences can be captured under the current experimental setting.

Table 1. Performance of CNN models in region classification tasks.

Models		Avg. Acc ± S.D. [a]	Avg. Pre ± S.D. [b]	Avg. Recall ± S.D. [c]	Avg. F1 ± S.D. [d]
EfficientNet_B1	Training set	0.9574 ± 0.051	0.9531 ± 0.050	0.9491 ± 0.053	0.9463 ± 0.052
	Validation set	0.9359 ± 0.038	0.9311 ± 0.023	0.9367 ± 0.037	0.9427 ± 0.032
	Test set	0.9081 ± 0.035	0.9048 ± 0.025	0.9193 ± 0.036	0.9210 ± 0.031
EffcientNet_B4	Training set	0.9483 ± 0.097	0.9496 ± 0.064	0.9486 ± 0.097	0.9225 ± 0.148
	Validation set	0.9104 ± 0.031	0.9261 ± 0.021	0.9104 ± 0.031	0.9052 ± 0.038
	Test set	0.8909 ± 0.047	0.8967 ± 0.047	0.8909 ± 0.047	0.8823 ± 0.062
MobileNet_v3_small	Training set	0.9287 ± 0.015	0.9259 ± 0.021	0.9387 ± 0.015	0.9057 ± 0.014
	Validation set	0.9102 ± 0.042	0.9308 ± 0.021	0.9019 ± 0.040	0.9022 ± 0.034
	Test set	0.9092 ± 0.036	0.9061 ± 0.021	0.9001 ± 0.034	0.9000 ± 0.024
MobileVit_xx_small	Training set	0.8919 ± 0.051	0.8737 ± 0.048	0.8593 ± 0.050	0.8975 ± 0.057
	Validation set	0.8871 ± 0.095	0.8698 ± 0.047	0.8787 ± 0.080	0.8803 ± 0.081
	Test set	0.8825 ± 0.067	0.8609 ± 0.036	0.8856 ± 0.055	0.8809 ± 0.059
MobileVit_x_small	Training set	0.8993 ± 0.047	0.8919 ± 0.036	0.8839 ± 0.047	0.8738 ± 0.053
	Validation set	0.8850 ± 0.064	0.8865 ± 0.026	0.8819 ± 0.052	0.8814 ± 0.053
	Test set	0.8834 ± 0.058	0.8892 ± 0.034	0.8758 ± 0.055	0.8760 ± 0.049

[a] Avg. Acc ± S.D. represents the mean accuracy ± standard deviation; [b] Avg. Pre ± S.D. represents the mean precision ± standard deviation; [c] Avg. Recall ± S.D. represents the mean Recall ± standard deviation; [d] Avg. F1 ± S.D. represents the mean F1 score ± standard deviation.

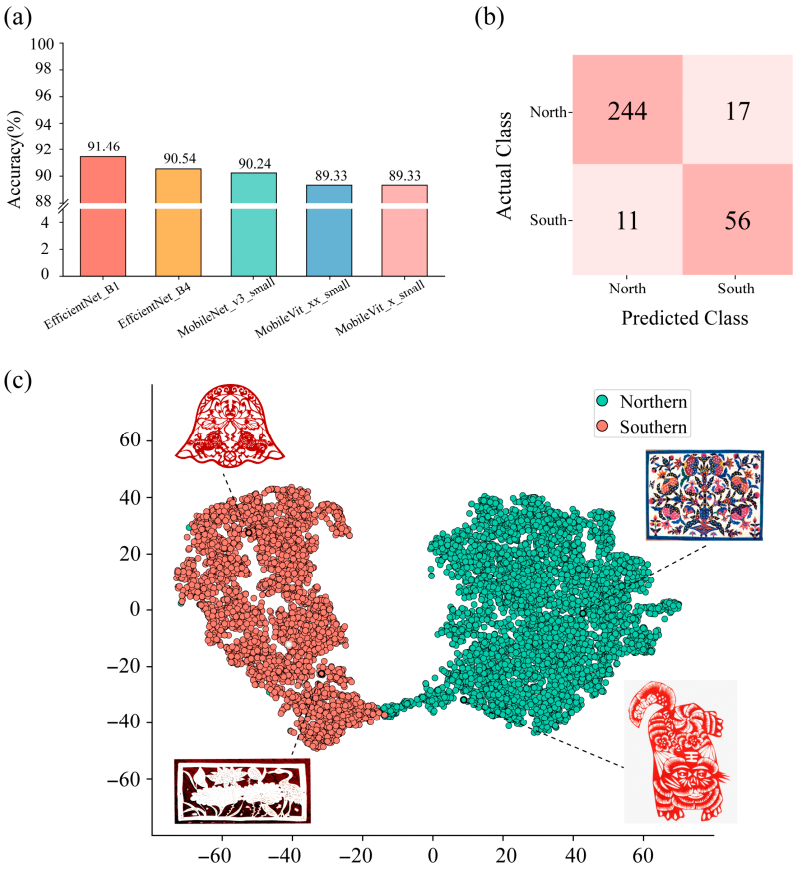


Figure 3. Performance analysis of CNN-based regional classification models for Chinese rural paper-cutting works. (**a**) Classification accuracy of five CNN models on the test set; (**b**) confusion matrix of the EfficientNet-B1 model; (**c**) dimensionality reduction of the penultimate layer features learned by EfficienNet_B1 based on t-SNE.

To further examine the classification behavior of the models, the confusion matrix of the EfficientNet-B1 predictions is shown in Figure 3b. The results demonstrate overall balanced performance between the two regional categories, indicating stable discriminative capability. In addition, t-distributed stochastic neighbor embedding (t-SNE) was

used to visualize the feature representations learned by EfficientNet-B1, as shown in Figure 3c. t-SNE is a widely used nonlinear dimensionality reduction technique designed to visualize high-dimensional data in two- or three-dimensional space [36]. The visualization shows that images belonging to the same regional category tend to form compact clusters, while samples from different regions are clearly separated. However, as a nonlinear and non-deterministic dimensionality reduction method, t-SNE is used here only for qualitative visualization.

3.2. Artistic Style Classification in Paper-Cutting Images

To assess model performance in the artistic style classification task for Chinese rural paper-cutting images, the classification results of the five CNN models were compared. For each model, we repeated the experiments using five different random seeds. Table 2 presents the performance for each class in terms of mean and standard deviation. The results show that although the performance varies across classes, the models maintain similar performance trends, indicating that they do not rely solely on the majority class. Figure 4a shows the overall best accuracy for each model. Among the evaluated architectures, MobileNetV3-Small achieved the best performance, with an accuracy of 73.20%. EfficientNet-B1 and EfficientNet-B4 followed closely, with accuracies of 72.46% and 71.71%, respectively. MobileViT-X-Small achieved an accuracy of 71.46%, while MobileViT-XX-Small showed the lowest accuracy at 68.49%. Overall, classification accuracy in the style task decreased by approximately 18–20 percentage points compared with the regional classification task. This reduction may be attributed to the fact that stylistic attributes in paper-cutting art (e.g., degrees of complexity, delicacy, or conciseness) are often subjective and exist along a continuous spectrum rather than as clearly separable categories. Moreover, individual paper-cutting works frequently exhibit hybrid or transitional stylistic characteristics, further increasing classification difficulty. In contrast, works from the same region tend to share more consistent compositional patterns and technical conventions, making them easier for CNN models to distinguish.

Table 2. Performance of CNN models in Artistic style classification tasks on test set.

Models	Category	Avg. Acc ± S.D. [a]	Avg. Pre ± S.D. [b]	Avg. Recall ± S.D. [c]	Avg. F1 ± S.D. [d]
EfficientNet_B1	Other	0.5730 ± 0.10	0.5910 ± 0.11	0.5139 ± 0.10	0.4894 ± 0.068
	Concise	0.5256 ± 0.076	0.5565 ± 0.036	0.4977 ± 0.091	0.5213 ± 0.057
	Rugged	0.5580 ± 0.043	0.5667 ± 0.033	0.5583 ± 0.049	0.5598 ± 0.017
	Intricate	0.6554 ± 0.060	0.6988 ± 0.087	0.6433 ± 0.053	0.6662 ± 0.040
	Delicate	0.7163 ± 0.022	0.7123 ± 0.027	0.7195 ± 0.051	0.7143 ± 0.020
EffcientNet_B4	Other	0.5389 ± 0.091	0.5779 ± 0.072	0.5722 ± 0.11	0.5651 ± 0.059
	Concise	0.5521 ± 0.075	0.5070 ± 0.064	0.4889 ± 0.072	0.5369 ± 0.058
	Rugged	0.5159 ± 0.075	0.5785 ± 0.094	0.5942 ± 0.058	0.5282 ± 0.052
	Intricate	0.6988 ± 0.065	0.7515 ± 0.057	0.6800 ± 0.095	0.7087 ± 0.044
	Delicate	0.7050 ± 0.021	0.7360 ± 0.021	0.7197 ± 0.021	0.7354 ± 0.016
MobileNet_v3_small	Other	0.5455 ± 0.14	0.5720 ± 0.13	0.5681 ± 0.14	0.5688 ± 0.14
	Concise	0.5720 ± 0.091	0.5004 ± 0.095	0.5713 ± 0.095	0.5829 ± 0.090
	Rugged	0.5414 ± 0.095	0.5208 ± 0.11	0.5423 ± 0.090	0.5259 ± 0.091
	Intricate	0.7056 ± 0.092	0.7032 ± 0.12	0.6800 ± 0.089	0.6834 ± 0.066
	Delicate	0.7223 ± 0.060	0.7261 ± 0.038	0.7144 ± 0.079	0.7187 ± 0.050
MobileVit_xx_small	Other	0.5587 ± 0.12	0.5288 ± 0.12	0.3472 ± 0.13	0.3530 ± 0.12
	Concise	0.5254 ± 0.15	0.5113 ± 0.055	0.4953 ± 0.18	0.5260 ± 0.17
	Rugged	0.5897 ± 0.16	0.5088 ± 0.16	0.4942 ± 0.16	0.4898 ± 0.16
	Intricate	0.6425 ± 0.081	0.6901 ± 0.079	0.6367 ± 0.14	0.6565 ± 0.040
	Delicate	0.6853 ± 0.031	0.6804 ± 0.024	0.6636 ± 0.036	0.7263 ± 0.026

Table 2. *Cont.*

Models	Category	Avg. Acc ± S.D. [a]	Avg. Pre ± S.D. [b]	Avg. Recall ± S.D. [c]	Avg. F1 ± S.D. [d]
MobileVit_x_small	Other	0.5621 ± 0.12	0.5344 ± 0.069	0.4819 ± 0.21	0.5452 ± 0.12
	Concise	0.5054 ± 0.09	0.5121 ± 0.057	0.5456 ± 0.12	0.4781 ± 0.085
	Rugged	0.5387 ± 0.0092	0.5956 ± 0.075	0.5109 ± 0.12	0.5322 ± 0.077
	Intricate	0.6203 ± 0.0070	0.5986 ± 0.083	0.6767 ± 0.085	0.6210 ± 0.051
	Delicate	0.7054 ± 0.0068	0.6807 ± 0.054	0.7439 ± 0.095	0.7107 ± 0.037

[a] Avg. Acc ± S.D. represents the mean accuracy ± standard deviation; [b] Avg. Pre ± S.D. represents the mean precision ± standard deviation; [c] Avg. Recall ± S.D. represents the mean Recall ± standard deviation; [d] Avg. F1 ± S.D. represents the mean F1 score ± standard deviation.

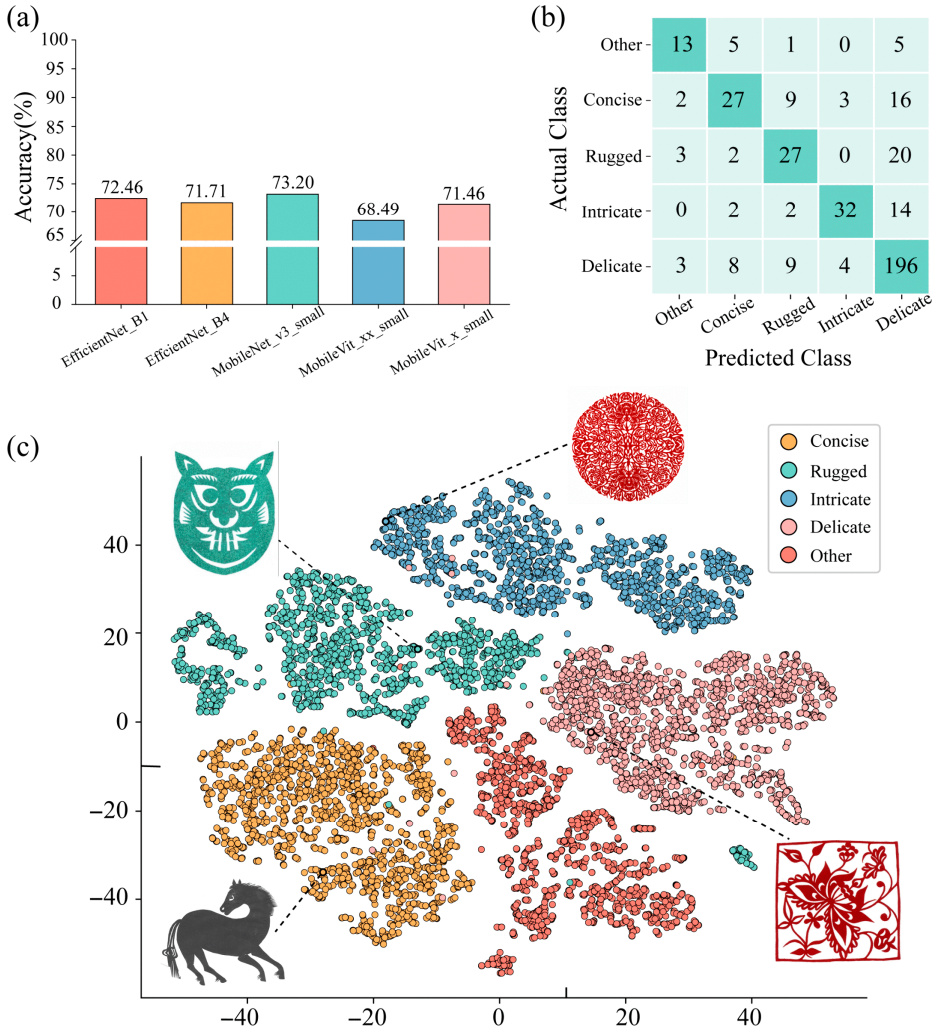


Figure 4. Performance analysis of CNN-based Artistic Style Classification models for Chinese rural paper-cutting works. (**a**) Classification accuracy of five CNN models on the test set; (**b**) confusion matrix of the MobileNet_v3_small model; (**c**) dimensionality reduction of the penultimate layer features learned by MobileNet_v3_small based on t-SNE.

The confusion matrix of the best-performing model, MobileNetV3-Small, is shown in Figure 4b. The results indicate that misclassifications primarily occur between visually adjacent styles, reflecting the inherent continuity and overlap among artistic styles in paper-cutting. Stylistic features often appear alternately contrasting and convergent rather than strictly distinct. To further explore the learned feature representations, t-SNE was applied to the high-dimensional features extracted by MobileNetV3-Small, as shown in Figure 4c. The visualization reveals that images of the same style tend to cluster in neighboring regions of the embedding space. Although some overlap between certain style categories is observed,

the majority of samples remain separable. As with the regional classification task, this visualization is intended for qualitative analysis only.

3.3. Interpretability Analysis

Deep learning models are often regarded as "black boxes" due to their high nonlinearity and multi-layered structures, making their internal decision-making mechanisms difficult to interpret directly [37]. In recent years, interpretability analysis methods have received increasing attention, as they help enhance the credibility of model outputs and provide insights into how models make decisions based on visual information [38]. In interdisciplinary contexts such as artistic image analysis, interpretability methods may offer useful perspectives for examining which visual regions contribute to model predictions, although such observations remain indirect.

In this study, Gradient-weighted Class Activation Mapping (Grad-CAM) was employed to conduct interpretability analysis on the trained deep learning models. Grad-CAM generates heatmaps that indicate model attention by leveraging the gradients of target classes with respect to the feature maps of the final convolutional layer, thereby revealing the image regions emphasized during classification decisions [39]. A key advantage of this method is that it does not require additional model training and can be applied directly to existing model weights.

EfficientNet-B1 and MobileNetV3-Small, the two best-performing models, were selected for visualizing decision-making processes in the regional classification and style classification tasks, respectively. In the generated heatmaps, colors ranging from blue to red indicate increasing levels of model attention, with red regions corresponding to visual features assigned higher importance during classification. It is worth noting that the examples selected in the following two subsections are intended to provide qualitative illustrations of model attention patterns rather than a comprehensive or statistically representative analysis.

3.3.1. Interpretability Analysis of Regional Classification

For the regional classification task, four representative paper-cutting artworks were randomly selected from the test set, including two from northern regions and two from southern regions. Figure 5 presents the Grad-CAM visualizations produced by the EfficientNet-B1 model for these samples.

In the northern paper-cutting examples, model attention is primarily concentrated on dominant subject forms with strong structural features and pronounced dynamic tension. For instance, in Figure 5a, a paper-cutting work themed around a vigorous tiger employs a diagonally oriented dynamic composition to convey a powerful and solid posture. The bold and forceful cutting techniques shape a full tiger head with round eyes and an expressive facial appearance, while the forward-striding forelimbs enhance the sense of motion. The heatmap indicates that the model focuses mainly on the tiger's head, body contours, and dynamic posture. In Figure 5b, the rounded tiger head is depicted roaring upward with wide-open eyes, while swirling decorative motifs amplify the overall sense of motion and tension, resulting in a fierce and vital visual impression. Attention is focused on the head region and surrounding decorative patterns. By contrast, in the southern paper-cutting examples, attention is more frequently distributed over fine-grained decorative elements and continuous line structures. In Figure 5c,d, the model highlights regions such as petals, stems, and flowing line patterns.

Overall, the Grad-CAM visualization results showcase the visual features that the model focuses on, which to some extent help assist human classification decisions and promote understanding of paper-cutting art.

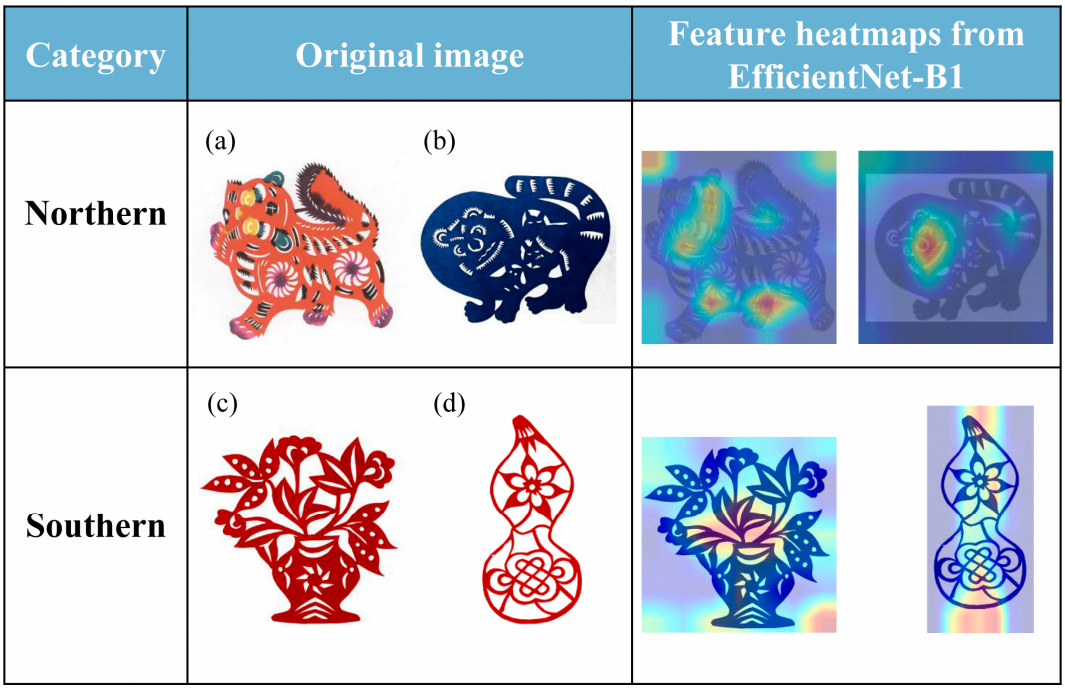


Figure 5. A comparison of the original paper-cutting images and corresponding feature heatmaps in a regional classification task using the EfficientNet-B1 model. (**a**,**b**) present representative samples from northern regions, while (**c**,**d**) correspond to representative samples from southern regions.

3.3.2. Interpretability Analysis of Art Style Classification

For the artistic style classification task, ten representative paper-cutting artworks were randomly selected from the test set, with two samples chosen for each style category. Figure 6 shows the Grad-CAM visualizations generated by the MobileNetV3-Small model for these samples.

In the rugged style samples (Figure 6a,b), the artworks typically employ bold cutting techniques and strong contrasts in form, emphasizing power and dynamic tension in the overall composition. In Figure 6a, an exaggerated and full tiger head combined with concise and forceful decorative patterns creates a vivid and expressive image with wide-open eyes. In Figure 6b, the classic scene of "Wu Song fighting the tiger" is rendered using large planar areas and strong lines, with the confrontation between human and animal reinforced through broad arcs and obtuse angular structures. The heatmaps of the two samples show that the model mainly focuses on the head, limb posture and their interrelationships of the figures and animals.

The delicate style samples (Figure 6c,d) exhibit highly dense and hierarchically organized visual structures. In Figure 6c, a multi-layered nested composition unfolds sequentially from mountains and flowing water to terraced fields, cave dwellings, and pastoral scenes, forming a visually rich yet orderly effect through dense line organization. In Figure 6d, a symmetrical folded peony motif serves as the central element, surrounded by expanding vine patterns that construct a stable and richly decorative structure. The visualizations indicate that the model focuses on high-density pattern regions and their repetitive structures, particularly areas of overlapping lines and concentrated details.

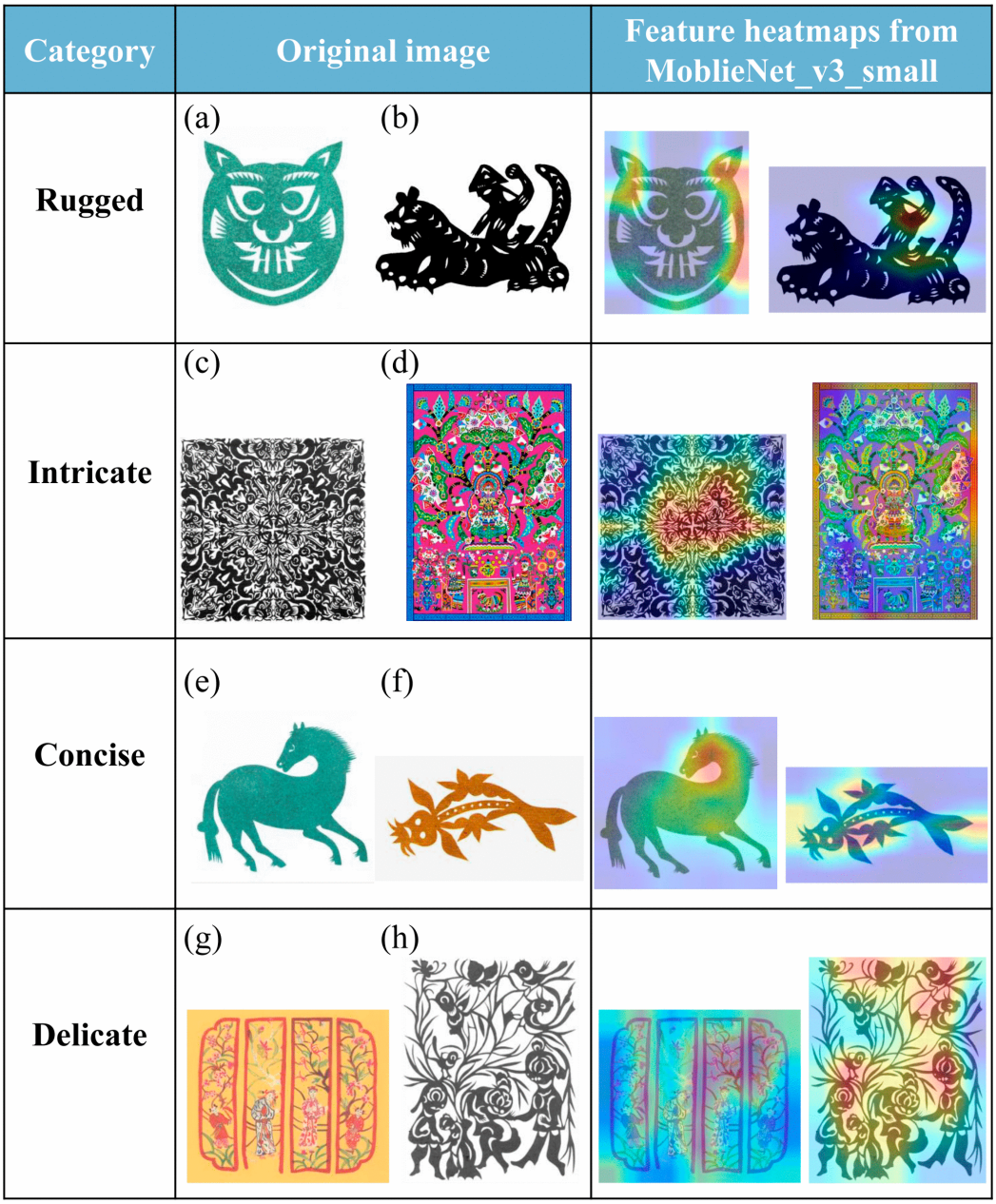


Figure 6. A comparison of the original paper-cutting image and the corresponding feature heatmap in a style classification task using the MobileNetV3-Small model. (**a**,**b**) present representative works of the rugged style, (**c**,**d**) present representative works of the intricate style, (**e**,**f**) present representative works of the concise style, and (**g**,**h**) present representative works of the delicate style.

In the concise style samples (Figure 6e,f), creators employ highly abstracted visual language to depict animal forms, conveying complete shapes and dynamic information with minimal lines and planar elements. The overall compositions are clear, with appropriate use of negative space to emphasize primary forms while downplaying secondary decorative details. The results show that the model's attention is concentrated on overall contours, head regions, and primary action areas, with relatively weak responses to backgrounds and minor details.

The intricate style samples (Figure 6g,h) reflect a high level of technical control and complex craftsmanship. In Figure 6g, multiple cutting techniques are combined to render figures and scenes with rich layering and decorative detail. In Figure 6h, dense and continuous curved lines organize the composition, interlinking human, animal, and plant

elements through fine linear connections to produce rhythm and order within a highly compact layout. The visualizations show that the model assigns high attention to fine structural elements, continuous curves, and intersections of decorative patterns.

Overall, the Grad-CAM visualizations reveal that the model tends to focus on visually salient regions that differ across samples, initially revealing the model's attention pattern.

4. Discussion

The contributions of this study are primarily twofold, encompassing data resource construction and methodological application. First, a large-scale image dataset of Chinese rural paper-cutting art covering multiple regions and styles was established. Against the backdrop of limited existing data resources, this dataset provides a systematic, standardized, and extensible foundation for future research and applications. It supports not only model training and performance evaluation in academic studies but also the organization and sharing of digital resources related to Chinese rural paper-cutting art. Second, from a methodological perspective, the CNN-based automated classification framework demonstrates strong practical value. The experimental results show that the models can predict the regional origin and artistic style of paper-cutting artworks with relatively high accuracy, offering a new pathway for computer-assisted analysis of traditional folk art images. In addition, interpretability analysis was employed to visualize the decision bases of the models. Compared with traditional classification approaches that rely on expert experience or subjective judgment, this method improves classification consistency and objectivity while enhancing transparency and credibility, providing an interpretable framework for digital classification and analysis of paper-cutting art.

To further verify the potential of this method in practical applications, we conducted a comparison with manual annotations. Taking the region classification task as an example, we randomly selected 100 images from the test set (50 from the southern region and 50 from the northern region) and compared the classification results of the best-performing convolutional neural network model, EfficientNet_B1, with the annotation results independently provided by three first-year university students who are beginners in paper-cutting art. As shown in Figure 7, compared with manual annotation, EfficientNet_B1 has significant improvements in both accuracy and annotation speed (only 6.3 s for 100 images with an accuracy of 91%). These findings indicate that the proposed method is not only efficient but also produces reliable and consistent prediction results.

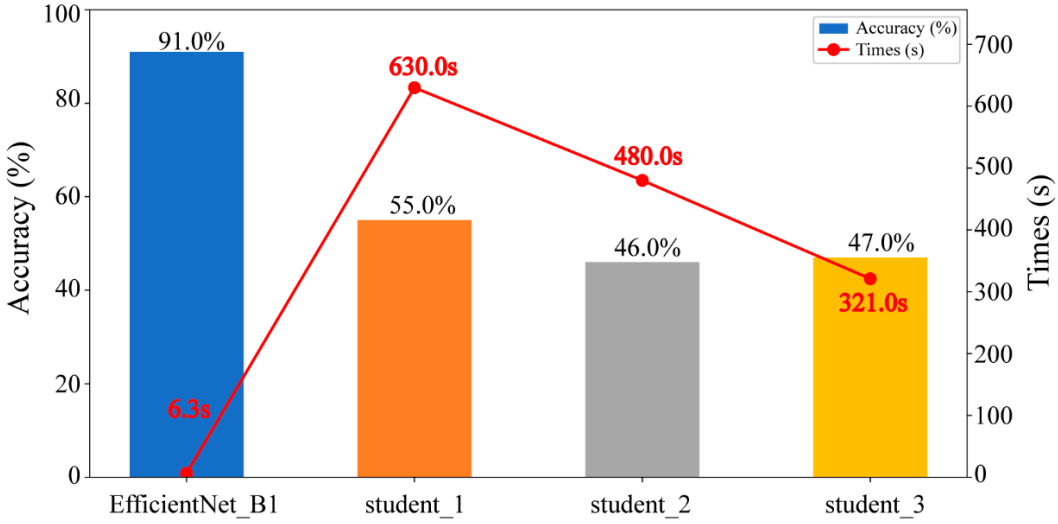


Figure 7. Performance of humans and CNN model in geographic classification task.

In future practical application scenarios, the proposed approach can be further integrated into digital exhibition platforms, educational support systems, and cultural heritage management systems to enable automatic annotation, classification-based retrieval, and intelligent recommendation of paper-cutting artworks. For example, in children's aesthetic education, teachers or students could photograph or upload paper-cutting works to quickly obtain information on regional cultural background and stylistic characteristics, offering immediate, intuitive, and accessible learning feedback. This process can help children understand aesthetic features and cultural meanings associated with different paper-cutting styles, thereby enhancing cultural understanding, artistic cognition, and independent aesthetic judgment. At the same time, the approach contributes to the popular dissemination and preservation of rural paper-cutting art, increasing the accessibility of this valuable cultural heritage. Future work will focus on expanding and refining the image database by incorporating works from additional regions, themes, and historical periods, thereby improving coverage of the overall spectrum of Chinese rural paper-cutting art and providing long-term support for digital preservation and cultural heritage protection.

5. Conclusions

This study investigated Chinese rural paper-cutting artworks by using a curated image collection encompassing multiple regions and artistic styles, and by applying five mainstream CNN models to classify regional and stylistic characteristics. In the regional classification task, all models achieved high accuracy, with EfficientNet-B1 performing best at 91.46%. The style classification task was more challenging, with MobileNetV3-Small achieving the highest accuracy of 73.20%. In addition, t-SNE visualizations provided a qualitative view of the learned feature distributions, suggesting a degree of separability between different categories. To further examine model behavior, Grad-CAM was used to visualize regions contributing to model predictions, offering preliminary insights into model attention patterns. Overall, this study demonstrates the effectiveness and feasibility of CNN-based approaches for analyzing Chinese rural paper-cutting art. These findings indicate potential applicability in areas such as digital organization, intelligent classification, and educational support for traditional folk art.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/app16073174/s1, Figure S1: The ROC curves for CNN models on training set, validation set, and test set in regional classification tasks. Figure S2: The ROC curves for CNN models on training set, validation set, and test set in art style classification tasks. Figure S3: The confusion matrices for CNN models on the training, validation and test sets in regional classification tasks. Figure S4: The confusion matrices for CNN models on the training, validation and test sets in art style classification tasks. Table S1: Hyperparameters of CNN models in regional classification tasks. Table S2: Hyperparameters of CNN models in art style classification tasks. Table S3. The performance of SVM using HOG features.

Author Contributions: Conceptualization, Y.X. and F.Z.; funding acquisition, Y.X.; investigation, X.W., X.Y. (Xiaoyue Yin), X.C. and X.Y. (Xudong You); validation, X.C. and X.Y. (Xudong You); writing—original draft, X.W. and X.Y. (Xiaoyue Yin); writing—review and editing, Y.X. and F.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Social Science Fund General Project for Education: "Research on a Model for Cultivating Artistic Creativity in Rural Children Using Convolutional Neural Network-Based Image Recognition" (Project Approval Number: BLA210218).

Data Availability Statement: The data presented in this study are available upon request from the corresponding author.

Acknowledgments: The authors would like to express their sincere gratitude to the following individuals: Zhang Xun, inheritor of Hunan Province's intangible cultural heritage of paper-cutting art; Kuang Xiaohua, inheritor of "Kuang's Paper-cutting"; and Zhang Xin and Wu Xuan, key rural art teachers. These paper-cutting inheritors and teachers generously provided a wealth of exquisite images of their collected paper-cutting works for our analysis and research. Their high-quality professional knowledge and selfless contributions have been crucial to the smooth progress of our research.

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

Abbreviations

The following abbreviations are used in this manuscript:

CNN	Convolutional neural networks
t-SNE	t-distributed stochastic neighbor embedding
Grad-CAM	Gradient-weighted Class Activation Mapping
ViTs	Vision transformers
ACC	Accuracy
ROC-AUC	Area under the receiver operating characteristic curve
Pre	Precision
TP	True positive
TN	True negative
FP	False positive
FN	False negative
TPR	True positive rate
FPR	False positive rate

References

1. Chen, T.; Wei, N. Analysis on the Visual Characteristics of the Traditional Patterns of Guangxin Paper Cutting. *Front. Art Res.* **2023**, *5*, 13–17. [CrossRef]
2. Guan, Y.; Pan, M. The Origin and Artistic Characteristics of Chinese Traditional Paper-Cut Art. In Proceedings of the 4th International Conference on Art Studies: Science, Experience, Education (ICASSEE 2020), Moscow, Russia, 27–28 August 2020; Atlantis Press: Cambridge, MA, USA, 2020. [CrossRef]
3. Chen, Y. The Application of Folk Paper-cut in Painting. *J. Humanit. Arts Soc. Sci.* **2023**, *7*, 664–667. [CrossRef]
4. Yang, X.; Yang, X. The Research Characteristics of Traditional Folk Art of Paper Cutting in Modern Society of China. In Proceedings of the 2019 International Conference on Arts, Management, Education and Innovation (ICAMEI 2019), Seoul, Republic of Korea, 12–14 January 2019; Clausius Scientific Press: Ottawa, ON, Canada, 2019. [CrossRef]
5. Shu, D.; Chen, Y. Research on the Integrity Protection of Intangible Cultural Heritage and Its Cultural Space—Take Paper-Cutting Art as an Example. *Cult. Int. J. Philos. Cult. Axiolog.* **2024**, *21*, 112–131.
6. Yang, Z.; Liu, Y.; Wang, W.; Li, Y.; Yuan, X. An empirical study of feature extraction for Hubei folk art of carved paper-cutting. *PLoS ONE* **2024**, *19*, e0311923. [CrossRef] [PubMed]
7. Gu, Y. Contrast and Fusion: The Role of Regional Culture in Shaping Jiangsu Paper-Cut Art. *J. Contemp. Educ. Res.* **2023**, *7*, 64–77. [CrossRef]
8. Wu, F. A Glimpse of Chinese Culture Through Papercuts. *Educ. Asia* **2004**, *9*, 40–46. [CrossRef]
9. Yu, J.; Chen, S. Research on Traditional Chinese Freehand Paper-cut. In Proceedings of the 2018 2nd International Conference on Education Science and Economic Management (ICESEM 2018), Xiamen, China, 25–26 August 2018; Atlantis Press: Paris, France, 2018. [CrossRef]
10. Chinese Paper-Cut—UNESCO Intangible Cultural Heritage. Available online: https://ich.unesco.org/en/RL/chinese-paper-cut-00219 (accessed on 23 January 2026).
11. Peng, D.; Sun, S.; Pan, L. Research on Chinese Paper-cut CAD System. In Proceedings of the Fourth International Conference on Image and Graphics (ICIG 2007), Chengdu, China, 22–24 August 2007; pp. 892–896. [CrossRef]
12. Faber, L.; Hartman, E.; Derikx, D.F.A.A.; Steggink, M.G.A.; Houwen, S.; Schoemaker, M.M. A Qualitative Observation Tool for Folding, Writing, and Cutting in School-Aged Children: Hands-On! *Can. J. Occup. Ther.* **2025**, *25*, 84174251397726. [CrossRef]
13. Lu, S.-Y.; Lo, C.-C.; Syu, J.-Y. Project-based learning oriented STEAM: The case of micro–bit paper-cutting lamp. *Int. J. Technol. Des. Educ.* **2022**, *32*, 2553–2575. [CrossRef]

14. Bohumelova, M.; Hvorecky, J. Digitized art promotes the cultural heritage. In *Proceedings of the 2015 13th International Conference on Emerging eLearning Technologies and Applications (ICETA), Starý Smokovec, High Tatras, Slovakia, 26–27 November 2015*; IEEE: New York, NY, USA, 2015; pp. 1–5. [CrossRef]
15. Chen, X.; Arya, V. Semantic Web Insights Into the Classification of Folk Paper-Cut Cultural Genes. *Int. J. Semant. Web Inf. Syst.* **2024**, *20*, 1–15.
16. 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]
17. Jogin, M.; Mohana Madhulika, M.S.; Divya, G.D.; Meghana, R.K.; Apoorva, S. Feature Extraction using Convolution Neural Networks (CNN) and Deep Learning. In Proceedings of the 2018 3rd IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT), Bangalore, India, 18–19 May 2018; pp. 2319–2323. [CrossRef]
18. Liang, S.; Pan, L.; Mohammadzadeh, F.; Khishe, M.; Ismail Kh, T.; Shambour, Q.Y.; Ghazal, T.M. Enhancing artistic style classification through a novel ArtFusionNet framework. *Sci. Rep.* **2025**, *15*, 36178. [CrossRef]
19. Cascone, L.; Nappi, M.; Narducci, F.; Russo, S.L. Classification of fragments: Recognition of artistic style. *J. Ambient. Intell. Humaniz. Comput.* **2023**, *14*, 4087–4097. [CrossRef]
20. Clay, V.; Schrumpf, J.; Tessenow, Y.; Leder, H.; Ansorge, U.; König, P. A quantitative analysis of the taxonomy of artistic styles. *J. Eye Mov. Res.* **2020**, *13*, 11. [CrossRef] [PubMed]
21. Li, W. Enhanced automated art curation using supervised modified CNN for art style classification. *Sci. Rep.* **2025**, *15*, 7319. [CrossRef] [PubMed]
22. Imran, S.; Naqvi, R.A.; Sajid, M.; Malik, T.S.; Ullah, S.; Moqurrab, S.A.; Yon, D.K. Artistic Style Recognition: Combining Deep and Shallow Neural Networks for Painting Classification. *Mathematics* **2023**, *11*, 4564. [CrossRef]
23. Feng, J. *Identifying Grassroots: A Classification Study of Chinese Folk Art*; Zhongzhou Ancient Books Publishing: Zhengzhou, China, 2006.
24. Sandoval, C.; Pirogova, E.; Lech, M. Two-Stage Deep Learning Approach to the Classification of Fine-Art Paintings. *IEEE Access* **2019**, *7*, 41770–41781. [CrossRef]
25. Jiao, L.; Zhao, J. A Survey on the New Generation of Deep Learning in Image Processing. *IEEE Access* **2019**, *7*, 172231–172263. [CrossRef]
26. Marques, G.; Agarwal, D.; De La Torre Díez, I. Automated medical diagnosis of COVID-19 through EfficientNet convolutional neural network. *Appl. Soft Comput.* **2020**, *96*, 106691. [CrossRef]
27. Tan, M.; Le, Q.V. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. *arXiv* **2020**, arXiv:1905.11946. [CrossRef]
28. Atila, Ü.; Uçar, M.; Akyol, K.; Uçar, E. Plant leaf disease classification using EfficientNet deep learning model. *Ecol. Inform.* **2021**, *61*, 101182. [CrossRef]
29. Dabwan, B.A.; Jadhav, M.E.; Olayah, F.A.; Shamsan, A.F. American gesture Language Detection Using Transfer Learning and Efficientnet B1 Technique. In *Proceedings of the 2023 International Conference on Recent Advances in Science and Engineering Technology (ICRASET), B G Nagara, India, 23–24 November 2023*; IEEE: New York, NY, USA, 2023; pp. 1–4. [CrossRef]
30. Howard, A.G.; Zhu, M.; Chen, B.; Kalenichenko, D.; Wang, W.; Weyand, T.; Andreetto, M.; Adam, H. MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications. *arXiv* **2017**, arXiv:1704.04861. [CrossRef]
31. Howard, A.; Sandler, M.; Chu, G.; Chen, L.-C.; Chen, B.; Tan, M.; Chu, G.; Vasudevan, V.; Zhu, Y.; Pang, R.; et al. Searching for MobileNetV3. *arXiv* **2019**, arXiv:1905.02244. [CrossRef]
32. Deng, L.; Du, Z.; Liu, X.; Wu, Z.; Lin, X.; Wen, B. Grape Leaf Cultivar Identification in Complex Backgrounds with an Improved MobileNetV3-Small Model. *Plants* **2025**, *14*, 3581. [CrossRef]
33. Ye, Y.; Hu, Z.; Diao, L.; Wang, Y. Grayscale Image Classification Method for Auto Parts Based on MobileNet_V3_Small. In *The 9th International Conference on Advances in Construction Machinery and Vehicle Engineering*; Saman, H., Peng, Y., Zhao, D., Bian, Y., Eds.; Springer Nature: Singapore, 2026; pp. 917–924. [CrossRef]
34. Mehta, S.; Rastegari, M. MobileViT: Light-weight, General-purpose, and Mobile-friendly Vision Transformer. *arXiv* **2022**, arXiv:2110.02178. [CrossRef]
35. Hsiao, S.-F.; Chao, T.-H.; Yuan, Y.-C.; Chen, K.-C. Hardware Accelerator for MobileViT Vision Transformer with Reconfigurable Computation. In Proceedings of the 2024 IEEE International Symposium on Circuits and Systems (ISCAS), Singapore, 19–22 May 2024; pp. 1–4. [CrossRef]
36. van der Maaten, L.; Hinton, G.; Rachmad, Y. Viualizing data using t-SNE. *J. Mach. Learn. Res.* **2008**, *9*, 2579–2605.
37. ŞAHiN, E.; Arslan, N.N.; Özdemir, D. Unlocking the black box: An in-depth review on interpretability, explainability, and reliability in deep learning. *Neural Comput. Appl.* **2025**, *37*, 859–965. [CrossRef]

38. Zhao, Z.-Q.; Zheng, P.; Xu, S.-T.; Wu, X. Object Detection With Deep Learning: A Review. *IEEE Trans. Neural Netw. Learn. Syst.* **2019**, *30*, 3212–3232. [CrossRef]
39. Selvaraju, R.R.; Cogswell, M.; Das, A.; Vedantam, R.; Parikh, D.; Batra, D. Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization. In Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017; pp. 618–626. [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.