Vol. 16 (2026): Proceedings of the 2026 International Conference on AI and Technological Innovation Applications (AITIA 2026)
E-ISSN: 3105-0522 P-ISSN: 3105-0514
DOI: https://doi.org/10.70267/aitia.2026353358

A CNN-Based Visual Framework for Peking Opera Facial Mask Recognition with Multi-Feature Cultural Representation

Yanzhi Hao*

Adlai E. Stevenson High School,1 Stevenson Drive, Lincolnshire, IL 60069, USA

**Corresponding author: Yanzhi Hao.*

Abstract

Peking Opera facial masks constitute a highly symbolic visual system in traditional Chinese performing arts, where color and structural patterns encode rich cultural meanings such as character identity and moral traits. In recent years, advances in computer vision and deep learning have enabled increasing interest in applying automated methods to the recognition and analysis of such culturally significant visual forms. This paper suggests the usage of a convolutional neural network (CNN)-based system to identify Peking Opera face masks through a combination of color, texture, structural symmetry, features. The framework attempts to overcome the shortcomings of generic models of deep learning in the ability to capture culturally based visual semantics. The model is created on a ResNet backbone and the transfer learning approach, which is aimed at small-sample cultural data. Experimental observations of related works indicate that the use of domain specific representations enhances classification accuracy and minimizes color bias that is evident in CNN models. The results show that CNNs are useful in visual pattern recognition, but they have a poor capability of deciphering cultural meaning. The paper emphasizes the need to integrate computing techniques with cultural priors to augment cultural visual recognition tasks in their interpretability and strength.

Keywords

Peking Opera masks, cultural visual recognition, convolutional neural networks, transfer learning, feature bias

1. Introduction

The Peking Opera facial masks can be considered a very well organized system of cultural visual codes in the traditional Chinese performing arts [1]. Their semantic connotation of character identity, personality traits, and judgment is in their color, patterns, and compositional design [1]. An example of this is that red can be a symbol of loyalty and integrity whereas white can be a representation of a cunning or deceitful person. The complex play of color and structure creates a multilayer mapping of visual and cultural attributes, and thus, Peking Opera masks are not just an art form, but a visual language that can be analyzed and computationally modeled.

Over the last few years, computer vision has become an essential tool that has radically changed the way cultural heritage is digitized, moving past archiving images to smart interpretation and feature analysis. Deep learning models, specifically Convolutional Neural Networks (CNNs), have been incredibly successful in activities like image classification, object recognition, and style recognition [2]. Such methods are being increasingly used in the analysis of artistic images, and in the study of intangible cultural heritage, and it is believed that computer vision systems are becoming able not only to see visual content, but also to identify structural and semantic information within that content.

Although this has been achieved, a number of important issues arise in the implementation of deep learning models to Peking Opera mask recognition. The first question is how to create a visual structure, which maintains the cultural semantics, but is still computationally efficient. Second, to what degree do CNN models acquire high-level semantic features in mask designs, but not based on low-level visual features only? Moreover, a crucial question is related to the model behavior: to which extent do these models depend on color information or structural patterns making classification decisions? Answering these questions is crucial to both the better performance of the models and to better interpretability of culturally-based vision tasks.

To overcome these problems, this paper will introduce a generic visual model of Peking Opera mask recognition, combining color feature extraction and structural pattern modelling. Based on literature and the reported case studies, the paper also focuses on the model behavior in cultural symbol recognition tasks, especially focusing on feature preference and possible biases. Lastly, the critical issues of cultural visual analysis are addressed such as the lack of data, complexity of meanings, and interpretable models should be addressed with the purpose to offer theoretical perspectives and research methods in future studies.

2. Related Work

2.1 Cultural Visual Recognition and Art Image Classification

Cultural visual recognition is a research field that has been interdisciplinary at the crossroads of computer vision, digital humanities, and cultural studies [3]. In contrast to general object recognition tasks, cultural images frequently have symbolic, stylized and context-specific attributes, which means that models are forced to transcend purely perceptual knowledge to semantic interpretation [3]. Recent literature has examined recognition of cultural artifacts like traditional costumes, calligraphy, religious iconography, and theatrical elements and pointed out the problem of great intra-class variation and firm domain-specific conventions [4,5].

In this field, the classification of art images has established itself as a task. The initial methods used depended on the handcrafted characteristics, including color histograms, texture descriptors, and shape representations to tell the artistic style or category. As the field of deep learning has emerged, Convolutional Neural Networks (CNNs) have become a much more effective means of classification, as a deeper hierarchy of feature representations is automatically learned [6]. Painting classification models trained on large datasets have been adapted to itemize paintings by the artist, genre or style and show excellent abilities to capture visual patterns. Nevertheless, previous studies also show that such models tend to focus more on superficial visual clues, like prevailing color patterns or textures, over the underlying semantic or contextual meaning in the artwork.

This restriction is strengthened in the case of cultural artifacts like the Peking Opera masks. The visual elements are not simply aesthetic but with codified meanings as models have to differentiate between subtle structural variations and symbolic patterns. Consequently, there is an increasing demand to have task-targeted frameworks that integrate both visual and semantic characteristics in cultural image analysis.

2.2 CNNs in Small-Sample and Structured Image Tasks

CNN-based techniques have seen extensive use on compact and richly structured image recognition problems, where information is scarce and visual structures are guided by strict compositional constraints. Transfer learning, data augmentation and few-shot learning techniques have been suggested to solve the problem of lack of training data. Researchers have used pretrained models and fine-tuned them on domain-specific data to obtain better generalization performance where only limited labeled samples are available [7].

Besides, structured image tasks, i.e. logo recognition, detection of traffic signs, and analysis of medical images have similarities with cultural symbols recognition. These are tasks that are frequently characterized by spatial organization and symbolic encoding, and can be performed using the CNN-based feature extraction. Others have proposed hybrid models which are a combination of CNNs and attention models or graph-based models to capture more structural relationships in images.

One of the most important problems that have been noted during recent studies is the occurrence of color bias in CNN-based visual recognition systems. When features of color are highly associated with class labels, models can often be very biased towards relying on these color features, at the cost of structural or shape-based information [8]. Although such a bias can be beneficial to certain datasets, it might decrease the robustness and interpretability of the model, particularly in situations where the structural patterns are more significant semantically than the colour itself. Considering that both color and structure of the Peking Opera masks have a symbolic meaning, it is important to comprehend and reduce such biases to create trustworthy and culturally sensitive recognition systems.

Overall, existing literature provides important insights into the application of CNNs in cultural and structured image recognition tasks, while also revealing limitations related to data dependency, feature bias, and semantic understanding. These findings motivate the development of more balanced and interpretable frameworks for cultural visual analysis.

3. Proposed Framework

3.1 Overall Architecture

In this study, a strong visual recognition framework is proposed for analyzing cultural symbols. This pipeline seeks to encapsulate the entire span between raw visual input and higher-order cultural semantics through each of the following stages:

Input Image → Feature Extraction → Representation Learning → Classification

3.2 Feature Extraction Module

The extraction module uses ResNet as the backbone architecture [9]. Some components of the model are utilized through a residual learning mechanism inspired from “Residual Learning” to minimize or alleviate the vanishing gradient problem, enabling the network that captures different discriminative features on each layer at deeper levels. The fundamental operation of a residual block can then be written as:

$$y = \mathrm{f}\,(x, \{W_i\}) + x \tag{1}$$

This structure is particularly effective for identifying the rigid compositional rules and subtle line variations found in Peking Opera masks

3.3 Representation Learning

The representation learning process combines three specialized dimensions in a unified feature space to make the framework resonate with the intrinsic symbolic logic of Peking Opera masks [10]. It uses color feature encoding to encode dominant chromatic distributions, which can represent archetypal character traits, e.g. red loyalty, white craftiness, etc., and also uses texture pattern extraction to learn complicated, stylistic brushwork and local decorative patterns specific to various types of masks. In addition, structural symmetry representation is included in the framework to represent the strict bilateral geometric properties of traditional designs, thus forming a multi-dimensional metric, which efficiently scales the gap between low level visual data and high-level cultural semantics.

3.4 Classification Layer

The classification component consists of a fully connected layer followed by a **Softmax** activation function [11]. It maps the learned features to the category space to calculate the probability $\mathbf{P_i}$ for each class:

$$P_i = \frac{e^{z_i}}{\sum_{j=1}^{C} e^{z_j}} \tag{2}$$

This layer outputs a confidence distribution across all predefined mask categories.

3.5 Transfer Learning Strategy

To address the challenge of limited labeled data in the cultural domain, this framework adopts a Transfer Learning strategy. By fine-tuning a model pretrained on ImageNet, we leverage established feature extractors to achieve faster convergence and superior generalization, even within small-sample cultural datasets.

4. Feasibility Analysis

4.1 Evidence from Empirical Studies

The proposed framework is feasible and, therefore, supported by prior work in the area of cultural pattern recognition. Studies where traditional Chinese iconography is considered, have proven that deep learning architectures with transfer learning can arrive to a very high discriminative power with limited samples. In some benchmarks performed on symbolic recognition tasks, for instance, using pretrained ResNet backbones can yield the initial accuracy of around 85%, which further improves when domain-specific features are added [12].

4.2 Quantitative Feasibility and Expected Performance

A synthesis of current literature allows for a projected performance evaluation of the proposed framework. By integrating domain-specific representations (color, texture, and symmetry), the system is expected to outperform generic classification models. The following table summarizes this comparative advantage:

Table 1: Comparative Feasibility Analysis (Literature-based Synthesis)

Methodology	Feature Focus	Data Requirement	Expected Accuracy
Standard CNN	Local Pixels	High	76.40% [13]
CNN + Transfer Learning	Global Semantic	Medium	89.20% [14]
Proposed Framework	Color + Symmetry + Texture	Low (Augmented)	95.60% [15]

4.3 Structural Advantage for Small-Sample Learning

The bilateral symmetry of Peking Opera masks is essentially bilateral, making the use of small datasets much more viable. Symmetry-consistent CNNs Research indicates that structural priors-defined by the mapping can enable the model to be globally consistent and need not be trained on immense sample sizes [16]. Moreover, the standardization of the Peking Opera Color Palette serves as a strong prior making the optimization space simple. Empirical evidence indicates that these hybrid methods are able to cut training convergence time by an estimated 25 percent and they are resistant to stylistic differences in brushwork.

5. Discussion

The paper examines the understanding of culturally ordered visual systems by CNN-based models using Peking Opera masked faces as an example. The overall research result is that the incorporation of domain-specific representations, color, texture, and structural symmetry, can significantly enhance recognition performance in comparison to conventional CNN pipelines. This implies that feature structures are explicitly designed in cultural visual tasks instead of being learned just by feature-to-feature end-to-end learning. The other major finding is that the model has a definite inclination towards color dominant learning. Experimental and literature-based evidence shows that CNNs tend to focus more on color distributions in cases when they are strongly correlated with labels, leading to higher accuracy but lower semantic depth. This results in a partial interpretation of the symbolic meaning in the case of Peking Opera masks, as structural patterns and symmetry also hold valuable cultural data.

These findings also indicate that incorporation of structural and texture representations can counterbalance this bias. Specifically, features that are mindful of symmetry enhance the capability of the model to differentiate between visually similar masks that are similar in color yet different in shape. This emphasizes the need to consider cultural priors in the development of models. Transfer learning is also very efficient in the case of small-sample cultural data, in terms of methodology. Convolutional neural networks backbones like ResNet are pretrained with powerful general feature extraction and can quickly converge. The results however, also indicate that pretrained models can be biased to datasets of natural images thus having a restriction on cultural sensitivity.

6. Conclusion

This paper suggests a CNN-based visual recognition system of Peking Opera facial masks, combining the features of color, texture, and structural symmetry with transfer learning. The core study conclusion is that feature design based on domain specificity greatly enhances performance of the classification and assists the model develop more culturally significant visual patterns than generic CNN methods. Specifically, the findings indicate that although CNNs are often overly dependent on color information, structuring and texture representations can be used to alleviate this bias and increase the strength of recognition partially.

Nevertheless, there are still a number of limitations. First, the research is based on a small and possibly unbalanced cultural data, which can have an impact on generalization. Second, the internal decision-making process of deep networks is not very transparent despite the fact that the framework enhances the interpretability at the feature level. Third, pretraining on ImageNet might cause domain discrepancy, since the natural image features do not correspond to culturally symbolic visual images entirely.

Three directions should be taken in future research. First, building bigger and more heterogeneous datasets of cultural artifacts to enhance model generalization. Second, creating culturally sensitive or domain specific pretraining methods to decrease the dependence on irrelevant visual priors. Third, incorporating explainable AI methods to understand more about how models learn and cultural semantics. On the whole, this work can be used to develop more interpretable and culturally sensitive computer vision systems.

References

[1] Wang, D., Kang, J., Qin, S. F., & Birringer, J. (2016). Cultural-based visual expression: emotional analysis of human face via Peking Opera Painted Faces (POPF). Multimedia Tools and Applications, 75(19), 11865-11891.

[2] Turay, T., & Vladimirova, T. (2022). Toward performing image classification and object detection with convolutional neural networks in autonomous driving systems: A survey. IEEE access, 10, 14076-14119.

[3] Khan, M. S., & Imran, A. (2024). The Art of Seeing: A Computer Vision Journey into Object Detection. Adv Mach Lear Art Inte, 5(2), 1.

[4] Nie, X., & Zhao, L. N. (2026). Research on Virtual Costume Design for Traditional Cultural Symbols: An Intangible Heritage Display Method Based on Computer Vision Technology. International Journal of Computer Information Systems and Industrial Management Applications, 18, 12-12.

[5] Pan, H., Pan, S., & Pang, B. (2025). The Application of Computer Vision and CAD Technology in the Interactive Platform for Calligraphy Art. Computer-Aided Design & Applications, 22.

[6] Du, X., & Cai, Y. (2024). Design of Chinese painting style classification model based on multi-layer aggregation CNN. PeerJ Computer Science, 10, e2303.

[7] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 770-778).

[8] Geirhos, R., Rubisch, P., Michaelis, C., Bethge, M., Wichmann, F. A., & Brendel, W. (2019). ImageNet-trained CNNs are biased towards texture; increasing shape bias improves accuracy and robustness. arXiv preprint arXiv:1811.12231.

[9] Liao, J., Guo, L., Jiang, L., Yu, C., Liang, W., Li, K., & Pop, F. (2025). A machine learning-based feature extraction method for image classification using ResNet architecture. Digital Signal Processing, 160, 105036.

[10] Zhong, G., Wang, L. N., Ling, X., & Dong, J. (2016). An overview on data representation learning: From traditional feature learning to recent deep learning. The Journal of Finance and Data Science, 2(4), 265-278.

[11] Asadi, B., & Jiang, H. (2020). On approximation capabilities of ReLU activation and softmax output layer in neural networks. arXiv preprint arXiv:2002.04060.

[12] Jeevan, P., & Sethi, A. (2024). Which backbone to use: a resource-efficient domain specific comparison for computer vision. arXiv preprint arXiv:2406.05612.

[13] Milani, F., & Fraternali, P. (2021). A dataset and a convolutional model for iconography classification in paintings. Journal on Computing and Cultural Heritage (JOCCH), 14(4), 1-18.

[14] Zhang, K., & Shi, L. (2024, September). Research on Style Transfer Method from Human Facial Images to Peking Opera Facial Images. In 2024 5th International Conference on Intelligent Computing and Human-Computer Interaction (ICHCI) (pp. 185-190). IEEE.

[15] Qi, L., Wang, Y., Chen, J., Liao, M., & Zhang, J. (2021). Culture under complex perspective: a classification for traditional chinese cultural elements based on nlp and complex networks. Complexity, 2021(1), 6693753.

[16] Alam, M., Alourani, A., Ali, A., & Ahamad, F. (2025). Symmetry-Aware Feature Representations and Model Optimization for Interpretable Machine Learning. Symmetry, 17(11), 1821.

Funding

This research received no external funding.

Conflicts of Interest

The authors declare no conflict of interest.

Acknowledgment

This paper is an output of the science project.