

<https://doi.org/10.1038/s40494-026-02413-x>

Research on Song dynasty copper mirror pattern recognition based on MOEAD

Qing Feng¹, Kexin Yu¹ ✉, Yaxuan Li¹ & Tian Ma²

Song-dynasty bronze-mirror pattern recognition has largely relied on subjective expert judgment, limiting both efficiency and accuracy. Here we develop an automated framework that integrates ResNet50 with the multi-objective evolutionary algorithm based on decomposition (MOEA/D) to identify animal motifs. We curate a dataset of bronze-mirror images covering 14 animal-pattern categories and use data augmentation to mitigate limited sample size and improve generalization. A ResNet50 classifier is trained and its hyperparameters are jointly optimized via MOEA/D, balancing performance objectives. The optimized model achieves a maximum Hamming accuracy of 94.48% on the validation set and shows improved predictive consistency. Macro-F1, which is sensitive to minority classes, varies modestly but overall generalization is strengthened. On the test set, the approach remains computationally tractable and outperforms comparative models. This multi-objective optimization strategy offers a robust route for automated identification and authentication of complex traditional decorative motifs, supporting broader dissemination and reuse in heritage-science workflows.

Bronze mirrors, a distinctive category of ancient Chinese bronzeware, were widely used in social life from the Shang and Zhou periods onward. They functioned not only as everyday practical objects but also as symbols of status, religious belief, and social aesthetics. Spanning multiple stages of Chinese arts and crafts history, bronze mirrors constitute important artistic and cultural media. Through social and cultural evolution across dynasties, they developed pronounced and diverse aesthetic styles. Specifically, the role of bronze mirrors in ancient Chinese society is reflected in several dimensions: first, as symbols of feudal authority; second, as symbols of culture and belief¹; third, as objects with practical and decorative functions; fourth, as instruments facilitating social interaction and emotional ties; fifth, as apotropaic devices used to ward off evil, including mirrors suspended on coffins². In addition, the reflective function of bronze mirrors was closely intertwined with socio-cultural practices and, particularly within religious rites and faith-based activities, was endowed with sacred meanings associated with expelling or restraining malevolent forces.

Over the course of historical development, bronze mirrors underwent marked stage-specific transformations. During the Spring and Autumn and Warring States periods, with the decline of the ritual–music system, bronze mirrors became increasingly widespread and their ornamentation grew more intricate, giving rise to motifs such as mountain patterns, coiled chi-dragon patterns, and linked-arc designs. By the Han and Tang dynasties, bronze mirrors reached a decorative zenith; inscribed mirrors, animal-pattern mirrors, and mirrors depicting mythological narratives became

prevalent. In the Sui–Tang era, motifs shifted from the Five Phases divine beasts toward more secular and life-oriented themes^{3,4}. And turning toward representations of everyday reality, thereby reflecting popular aspirations for a better life⁵.

By the Song dynasty, rapid commercial expansion and the prosperity of handicraft industries fostered the rise of an urban citizen class dominated by merchant and artisan groups. In the Southern Song, a new landscape emerged through the integration of northern and southern craft techniques. A large number of mirror-production workshops appeared, and bronze-mirror motifs increasingly adopted standardized formats. Alongside the growth of literati culture and technological advances, the decorative styles of Song mirrors became conspicuously more diverse. Their subjects broadened to include floral and vegetal patterns, animal motifs, and inscriptions, gradually forming distinct period characteristics. Meanwhile, market mechanisms for bronze mirrors became more complex and mature, and mirrors were widely exported to Japan, Korea, Inner Mongolia, and other regions, serving as important media of cross-cultural exchange. The carving style of Song mirrors shifted from the fine, shallow relief typical of the Tang to flatter carving techniques, with denser and more meticulous compositional layouts.

At present, large numbers of Song bronze mirrors have been excavated, and some studies have employed 3D modeling techniques to restore Song mirrors and other cultural-heritage objects⁶. The importance of Song bronze-mirror art within the history of arts and crafts lies not only in its

¹School of Art, Xi'an University of Science and Technology, Xi'an, Shaanxi, China. ²College of Artificial intelligence & Computer Science (College of Software), Xi'an University of Science and Technology, Xi'an, Shaanxi, China. ✉e-mail: 1971478692@qq.com

inheritance of earlier mirror types, but also in innovations in surface ornamentation, casting technology, and cultural symbolism⁷. Scholars have further proposed novel approaches, such as the “gene-mapping” method, to analyse internal structures and evolutionary logics of traditional crafts⁸. However, existing research on Song bronze mirrors remains largely focused on reconstructing historical contexts, offering qualitative accounts of cultural value, and discussing casting techniques; most work is qualitative, and current methods show clear limitations.

To move beyond predominantly qualitative descriptions, recent work has increasingly explored computational approaches for motif visibility enhancement, recognition, and interpretation in cultural-heritage imagery. At the image level, fusion and enhancement methods have been used to improve the legibility of fine decorative structures and defect-related details, providing more reliable inputs for subsequent analysis⁹. Learning-based pipelines further extend this direction to automatic segmentation and motif recognition under degradation, such as corrosion and abrasion¹⁰. While evolutionary architecture search, feature-fusion strategies, and detection frameworks have also been reported for related heritage-pattern tasks¹¹. In parallel, lightweight network design supports mobile and embedded deployment for in situ inspection¹², and semantic knowledge modeling (e.g., knowledge graphs and QA systems) complements visual recognition by linking motifs to crafts and cultural meanings^{13,14}. Methodological surveys likewise highlight a broader shift from handcrafted descriptors to representation learning for detection and segmentation in heritage contexts¹⁵. Taken together, the literature is converging toward multimodal fusion, deployable real-time models, and integrated visual-semantic pipelines^{9–16}.

Within these pipelines, CNNs are a widely adopted backbone because they learn hierarchical multi-scale representations—from local texture cues to higher-level semantic abstractions—without extensive manual feature engineering^{17,18}. Representative studies have augmented CNNs to better capture multi-scale structure and improve robustness, including wavelet-domain feature mapping and frequency-domain residual fusion in a dual-task framework¹⁹, differential feature fusion within residual blocks to strengthen feature reuse and semantic continuity²⁰. And attention-augmented residual architectures for fine-grained style-related feature extraction²¹. Complementary efforts address computational efficiency (e.g., reducing convolutional energy cost)²² and analyse how architectural components such as kernel scale, residual connections, and normalization affect texture encoding and semantic representation^{23,24}. Overall, these advances suggest a progression from single-scale feature extraction to cross-layer and cross-scale fusion, providing a technical basis for high-precision recognition and subsequent semantic modeling in heritage imagery.

However, a persistent challenge in practical settings—especially under small, long-tailed, and noisy datasets—is to jointly balance recognition accuracy, model complexity, and generalization. Conventional single-objective optimization used in many classification or clustering pipelines often struggles to negotiate these competing criteria. Multi-objective optimization provides a principled alternative by optimizing multiple conflicting objectives simultaneously, thereby offering a mechanism for model-structure adjustment, feature weighting, and performance improvement in image recognition. The decomposition-based multi-objective evolutionary algorithm MOEA/D proposed by ref. 25 decomposes a multi-objective problem into a set of scalar subproblems and coordinates cooperative search via weight vectors, making it a natural candidate for multi-criteria optimization in recognition settings.

In practical image-recognition applications, MOEAs have been shown to improve classification accuracy and feature separability. Luo et al. proposed the MOASCC (Multi-objective Adaptive Simultaneous Clustering and Classification) algorithm, which treats fuzzy-clustering connectivity and classification error rate as dual objective functions, and introduces a Bayesian feedback mechanism to enable synchronous learning of clustering and classification²⁶. In texture-image and synthetic aperture radar (SAR) segmentation experiments, MOASCC achieved

higher classification accuracy and more stable convergence than MSCC, SVM, and RBFNN, demonstrating the advantages of multi-objective optimization for learning complex image features. In addition, Zhang et al. developed a global–local cooperative MOEA/D (GL-MOEA/D)²⁷, which incorporates local neighborhood search and global crossover strategies to rapidly approximate the Pareto front while maintaining a balanced distribution of solutions. Its performance in feature-weight selection and model-parameter optimization indicates that the MOEA/D framework can effectively control model complexity without sacrificing classification performance. Decomposition-based MOEAs such as MOEA/D, by explicitly trading off accuracy against complexity and global exploration against local convergence, exhibit strong adaptive optimization capacity in image-recognition tasks. Their cooperative mechanisms allow high-discriminability feature subspaces to be automatically identified while preserving model generalization, making them reliable tools for multi-objective recognition models built on deep features. Building on this theoretical framework, the present study introduces MOEA/D as the core algorithm for feature selection and parameter optimization to improve convergence behavior and overall recognition performance. Although MOEA/D has advantages in feature optimization, its application in the cultural-heritage domain remains limited. We therefore combine ResNet50 with MOEA/D to address the recognition of complex decorative motifs.

Specifically, the main objectives of this research are as follows.

1. To improve the accuracy and efficiency of bronze-mirror motif recognition. By integrating the multi-objective evolutionary algorithm (MOEA/D) with a ResNet50 model, we collaboratively optimize model hyperparameters to better accommodate the complexity and diversity of bronze-mirror motifs, thereby enhancing automated recognition performance.
2. To advance the digital preservation of bronze-mirror motifs. Using computer-vision methods and deep-learning models, we enable digital analysis and recognition of Song-dynasty bronze-mirror patterns, providing technical support for the digital archiving and conservation of cultural heritage.
3. To reduce subjectivity and error inherent to traditional methods. By adopting a systematic, automated deep-learning framework, we minimize manual intervention and associated biases, improving objectivity and reliability, and further strengthening automation and precision in cultural-heritage research.

Overall, this study not only promotes the development of automated identification in Song-dynasty bronze-mirror scholarship, but also offers a theoretical basis and practical reference for the systematic analysis of complex traditional motifs.

To clarify the research logic and establish evaluation criteria, we propose the following research hypotheses and quantitative targets based on the problem statement and study objectives.

Research hypotheses 1: Introducing the decomposition-based multi-objective evolutionary algorithm (MOEA/D) improves the performance of a ResNet50 model for recognizing Song-dynasty bronze-mirror motifs.

Research hypotheses 2: Under the joint objectives of minimizing validation loss while maximizing Hamming accuracy and Macro-F1, the multi-objective optimization algorithm can maintain training stability, thereby enhancing overall generalization.

Quantitative objectives 1: On the test set, the model achieves an increase in Hamming accuracy of no less than 5% relative to the baseline ResNet50.

Quantitative objectives 2: On the validation set, the model attains a Macro-F1 score above 0.60.

Quantitative objectives 3: The validation loss is stably controlled below 0.12.

Quantitative objectives 4: Under multi-class imbalanced sampling, the variance of F1 scores across categories remains within a difference of 0.25, thereby verifying recognition stability and robustness.

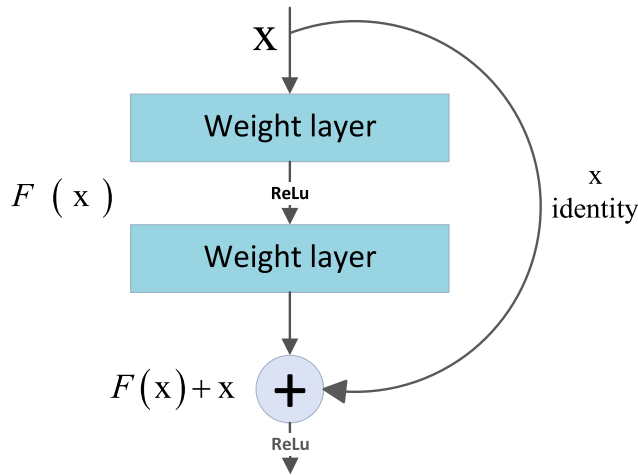


Fig. 1 | Residual structure of ResNet50. The input features pass through two network layers and are then directly added to the original input to form the output, thereby enhancing feature propagation and improving the training performance of deep neural networks.

Methods

Existing studies on bronze-mirror motif recognition have relied on conventional feature-extraction techniques and deep-learning approaches, yet there remains substantial room for improving recognition accuracy. To address these limitations, we propose a hybrid optimization framework that integrates a multi-objective evolutionary algorithm (MOEA/D) with a residual network (ResNet50) to identify animal motifs on Song-dynasty bronze mirrors. Specifically, MOEA/D is employed to optimize key network hyperparameters—including the learning rate and weight decay—thereby enhancing motif-recognition accuracy. ResNet50 offers a favorable balance between network depth and computational complexity, making it particularly well suited to our dataset scale and computational-resource constraints.

Deep neural network

To effectively extract complex decorative features from Song-dynasty bronze-mirror images and perform multi-label classification, we adopt ResNet50 as the backbone for deep feature extraction. ResNet50 is a classical variant of the residual network (ResNet) family. By introducing shortcut (residual) connections, it mitigates vanishing-gradient and model-degradation issues in deep neural networks, thereby improving both performance and training efficiency. Unlike traditional convolutional neural networks (CNNs), which are constructed by stacking layers sequentially, ResNet adds skip connections within each residual unit. This design allows the model to access the untransformed input from the preceding layer directly when learning feature transformations, reducing information loss and computational burden while enabling residual mapping. The mechanism can be expressed as in Eq. (1):

$$X_{l+1} = F(X_l) + X_l = F(X_l) + W_s \cdot X_l \tag{1}$$

Here, X_l denotes the input feature map at layer l ; $F(X_l)$ is the residual transformation function, typically implemented by two or three convolutional layers; and W_s is a projection matrix used for dimensional alignment. When the input and output dimensions are identical, the skip connection performs element-wise addition directly. This design enables the network to explicitly capture structural discrepancies between input and output representations, thereby improving learning efficiency. The residual architecture of ResNet50 is illustrated in Fig. 1. Within each residual block, skip connections transmit input information directly to the output, alleviating vanishing-gradient issues during deep-network training and enhancing performance in deep image-recognition tasks. Through

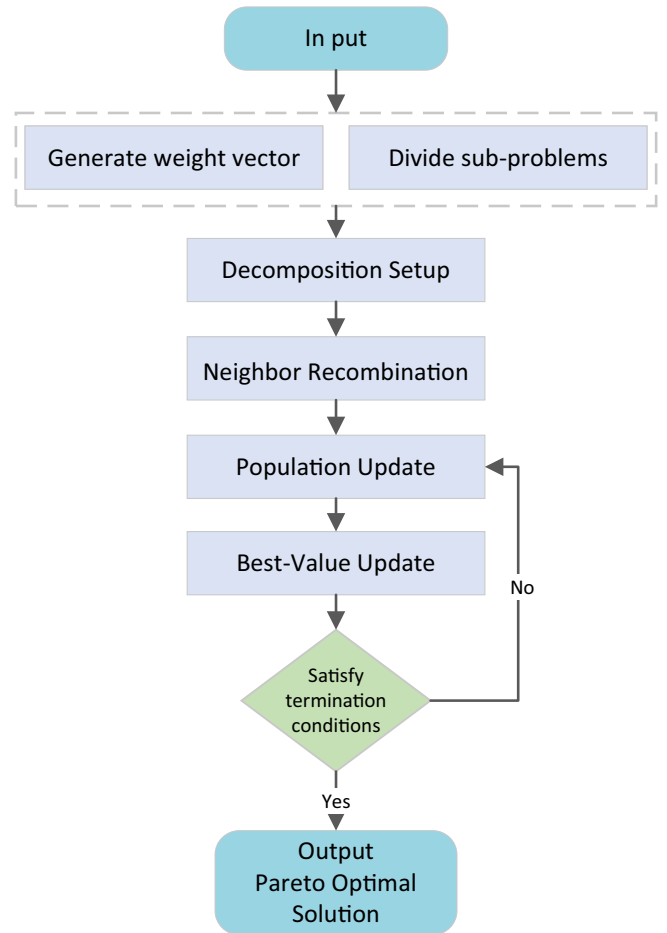


Fig. 2 | The process of selecting the optimal solution in MOEA/D. Firstly, the weight vector is generated and the problem is decomposed by dividing it into sub-problems. Subsequently, through neighborhood reorganization, population update, and optimal value update, iterative search is conducted continuously. Finally, after meeting the termination conditions, the Pareto optimal solution is output.

hierarchical feature learning—from shallow edge and geometric cues to deeper semantic and symbolic compositions—this architecture effectively extracts multi-level characteristics of Song-dynasty bronze-mirror motifs, providing a robust foundation for subsequent recognition and classification.

Multi-objective evolutionary algorithm based on decomposition

Compared with conventional grid search or random search, MOEA/D is not a passive “trial–evaluation” procedure. Instead, it performs an evolutionary search that actively exploits historical information to refine the solution set. Its decomposition strategy allows each subproblem to conduct fine-grained parameter exploration within a local region while preserving guidance toward the global optimum, giving it clear advantages when the search space is large and objective conflicts are pronounced. To address the multi-objective nature of CNN architecture optimization, we introduce the Multi-Objective Evolutionary Algorithm based on Decomposition (MOEA/D). As shown in Fig. 2, MOEA/D does not treat a multi-objective problem as a single monolithic task; rather, it uses weighted decomposition to transform it into a collection of single-objective subproblems that are solved cooperatively. In each generation, MOEA/D produces new candidate solutions and improves the solution for each subproblem via crossover and mutation operations. Through iterative evolution, the model hyperparameters are continuously adjusted, ultimately yielding higher accuracy and lower loss on both the training and test sets. Formally, the multi-objective optimization

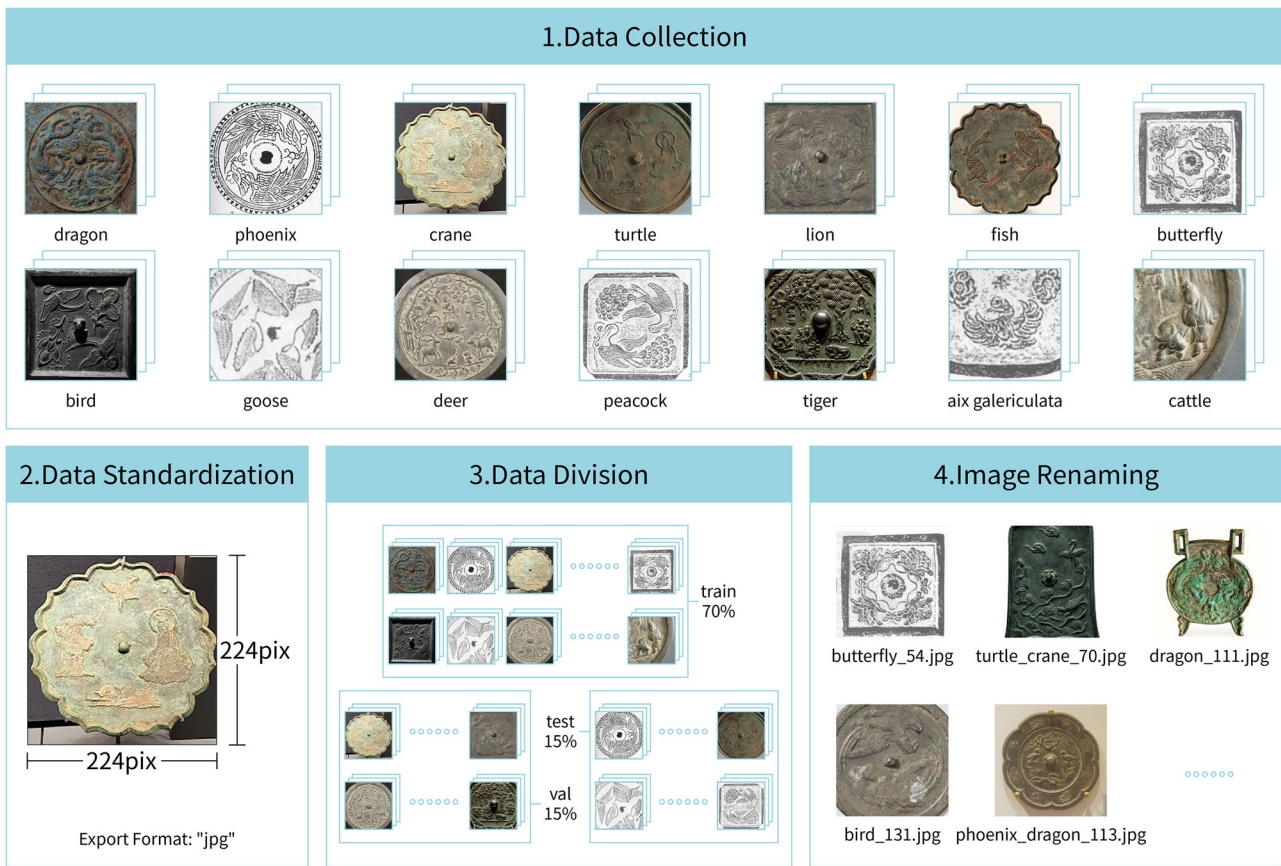


Fig. 3 | Dataset preprocessing and curation workflow. This figure illustrates the dataset preprocessing and curation workflow, which consists of four main steps: data collection, data standardization, data division, and image renaming.

problem in MOEA/D can be expressed as Eq. (2):

$$\text{minimize } F(X) = (f_1(X), f_2(X), \dots, f_M(X)), \tag{2}$$

$$\text{s.t. } X \in \Omega$$

Here, $X = (x_1, x_2, \dots, x_m)$ is a solution vector composed of N decision variables; $F(X)$ denotes an objective vector consisting of M objective functions $f_1(X), f_2(X), \dots, f_M(X)$ and the feasible solution space is Ω .

With a set of uniformly distributed weight vectors W_1, W_2, \dots, W_{NP} (where $W_i = (W_i^1, W_i^2, \dots, W_i^M)$), MOEA/D decomposes the original multi-objective problem into a series of weighted subproblems, formulated as Eq. (3):

$$\text{minimize } g_{ws}(X|W_i) = \sum_{j=1}^M w_i^j \cdot f_j(X) \tag{3}$$

This decomposition strategy enables the algorithm to optimize multiple objectives simultaneously and to share information across subproblems through a neighborhood search scheme, thereby approaching the Pareto-optimal set. The neighborhood of each subproblem is constructed according to the Euclidean distances between weight vectors. For each subproblem i , the $T = 5$ weight vectors with the smallest distances to its own weight vector are selected to form its neighborhood $B(i)$. This neighborhood structure constrains the local scope of cooperation during crossover and update operations and constitutes a key mechanism of MOEA/D. Crossover is implemented deterministically (crossover rate = 1.0), meaning that crossover is applied to parent individuals in every generation to produce

new solutions. The expansion factor of the crossover operator is set to $\gamma = 0.5$ to control the magnitude of offspring deviation.

Integrating MOEA/D optimization with CNNs

ResNet50 has been widely applied across domains such as image classification and object detection, and its effectiveness has been extensively validated and refined. Combining ResNet50 with the multi-objective evolutionary algorithm based on decomposition can further improve both feature extraction and hyperparameter tuning. In particular, MOEA/D optimizes network hyperparameters to enhance the accuracy and efficiency of ResNet50 for motif-recognition tasks, especially when dealing with multi-objective settings and complex patterned images. In this study, MOEA/D is configured to maximize training accuracy and Macro-F1 while minimizing test loss. The optimization focuses on two key hyperparameters: the learning rate (lr) and weight decay (wd). Each hyperparameter varies within predefined upper and lower bounds, and the evolutionary process dynamically adjusts their combinations to minimize the objective functions. The selection of each decision variable exerts a substantial influence on overall model performance.

Results

Dataset construction

To evaluate the performance of the ResNet50 model optimized by MOEA/D for recognizing decorative motifs on Song-dynasty bronze mirrors, we constructed a dedicated image dataset with animal motifs as the target categories. The dataset contains 140 original images of Song bronze mirrors, covering 14 motif classes: dragon, phoenix, crane, tortoise, lion, fish, butterfly, bird, wild goose, deer, peacock, tiger, mandarin duck, and ox. The procedures for image acquisition, processing, and categorization are shown

in Fig. 3. All images were resized to a unified resolution and manually annotated and categorized according to mirror names and motif classes. The full dataset was then split into training, validation, and test sets at a ratio of 70, 15, and 15%, respectively; class distributions across the three subsets

Table 1 | Distribution of the 14 motif categories in the original dataset, excluding augmented samples

Class	Training set	Test set	Validation set
<i>Aix galericulata</i>	2	1	1
Bird	6	2	2
Butterfly	4	1	1
Cattle	2	1	1
Crane	18	3	8
Deer	1	1	2
Dragon	25	5	3
Fish	16	3	1
Goose	5	1	1
Lion	1	1	0
Peacock	1	1	1
Phoenix	23	2	4
Tiger	1	1	0
Turtle	15	4	4

are reported in Table 1. All images were normalized using the ImageNet standard mean and variance. The dataset was expanded from publicly available image collections, with broad and heterogeneous sources, leading to partial inter-class overlap and an imbalanced class distribution. To improve model generalization, online data augmentation was applied to the training set, with each image randomly augmented 15 times (Fig. 4). The augmentation operations and their parameters are as follows:

1. RandomResizedCrop: Images are randomly cropped and resized to 224×224 pixels. The cropped area is randomly selected between 20 and 100% of the original image, allowing the model to adapt to objects at different scales.
2. RandomHorizontalFlip: Each image has a 50% probability of being horizontally flipped.
3. RandomRotation: Images are randomly rotated within a range of -15° to $+15^\circ$.
4. ColorJitter: Brightness, contrast, saturation, and hue are randomly perturbed, with ranges of ± 0.2 for brightness, ± 0.2 for contrast, ± 0.2 for saturation, and ± 0.1 for hue.
5. RandomAffine: Affine transformations are applied with rotation angles in -15° to $+15^\circ$, translations up to 10% of image width and height (translate = (0.1, 0.1)), and scaling between 0.9 and 1.1 (scale = (0.9, 1.1)).
6. RandomErasing: Each image has a 50% probability of undergoing random erasure. The erased area covers 2 to 25% of the original image (scale = (0.02, 0.25)), with an aspect ratio randomly drawn between 0.3 and 3.3 (ratio = (0.3, 3.3)).

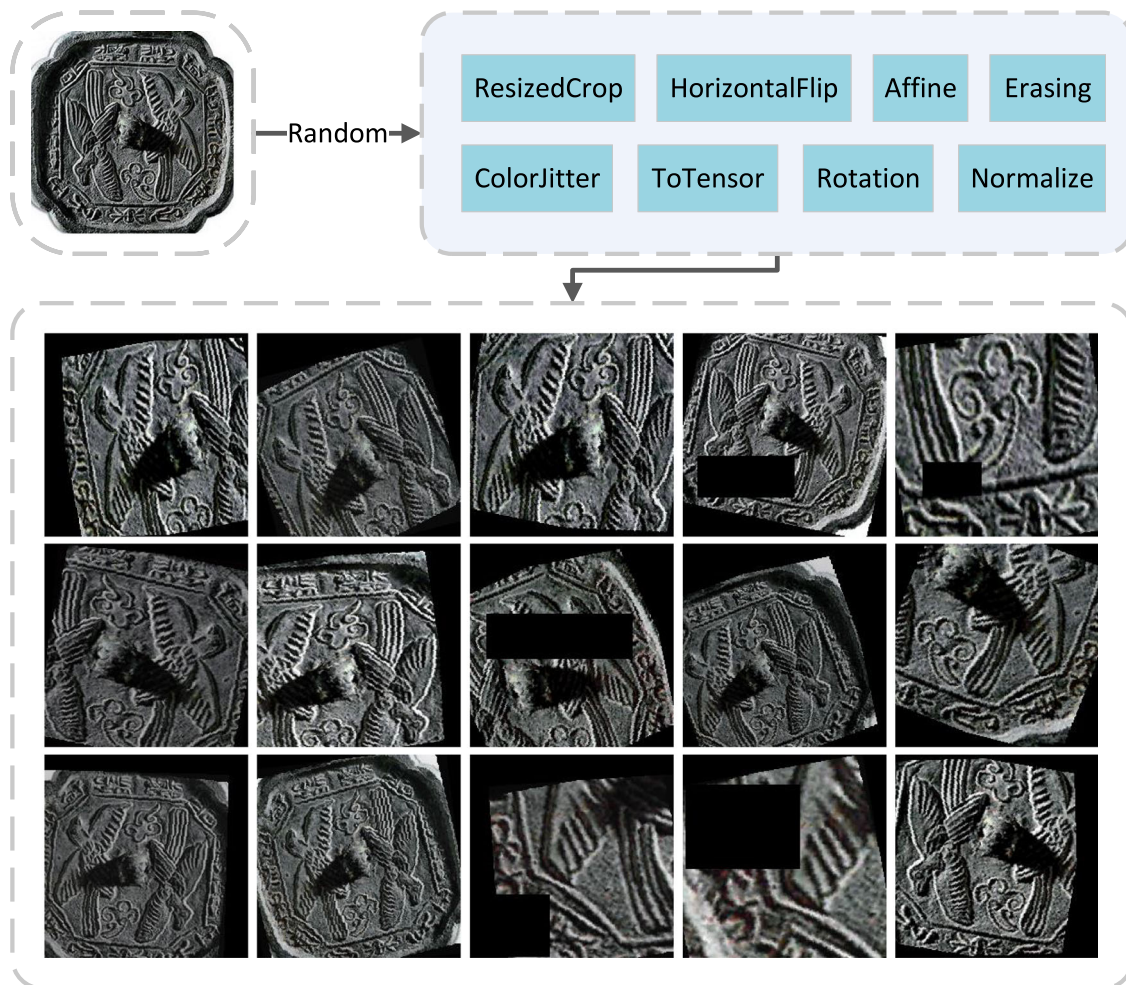


Fig. 4 | Examples of random image data augmentation. The top panel shows the original image and candidate augmentation operations; the bottom panel shows representative augmented results. The arrow indicates the augmentation process.

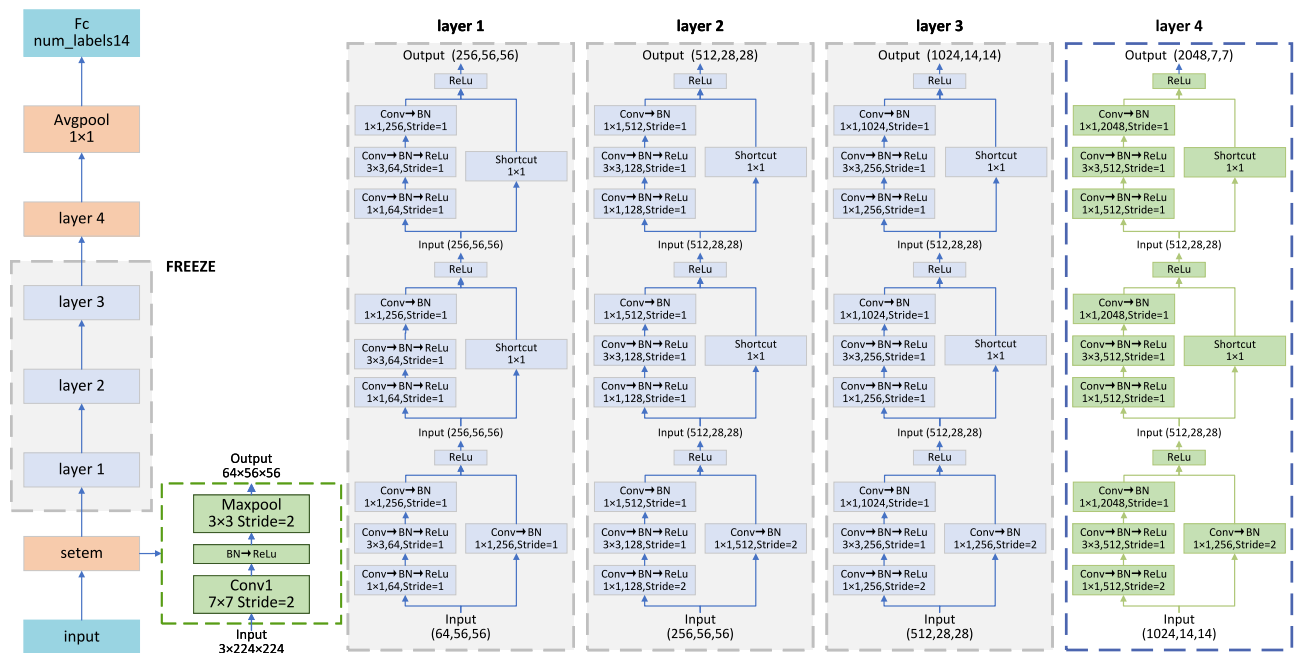


Fig. 5 | ResNet50 model architecture diagram. The gray background areas are all frozen regions and are not involved in later training, only using pretrained values.

7. ToTensor and Normalize: Images are converted to tensors and normalized using ImageNet statistics, with mean [0.485, 0.456, 0.406] and standard deviation [0.229, 0.224, 0.225].

Figure 4 illustrates the effects of these augmentation operations, including random cropping, color variation, rotation, flipping, and erasing. By exposing the model to more diverse visual instances during training, these strategies enhance robustness to real-world image variability and reduce the risk of overfitting to a narrow training distribution.

Experimental settings

We implemented a ResNet50-based multi-label image-classification model using the PyTorch framework, and performed performance optimization with MOEA/D in MATLAB. The backbone network was initialized with a pretrained ResNet50. As illustrated in Fig. 5, ResNet50 consists of an initial convolutional stem, four stages of residual blocks arranged as 3x, 4x, 6x, and 3x, and a final fully connected output layer that produces a 14-dimensional probability vector for multi-label prediction.

Within each Bottleneck residual block, the main branch first applies a 1x1 convolution for dimensionality reduction, compressing the channel number to planes. A 3x3 convolution is then used for spatial feature extraction at the reduced channel width, with the first block of certain stages employing stride 2 for downsampling. Finally, another 1x1 convolution expands the channels back to 4x planes. In the shortcut branch, when the input and output differ in channel number or spatial resolution, a 1x1 convolution (optionally with stride 2) is applied to match the bypass to the main-branch output shape; the two paths are then summed element-wise, forming the residual connection. The early layers of ResNet50 were frozen to retain pretrained representations, and only the final layers were trained.

In this experimental setup, key hyperparameters of ResNet50 were optimized, with MOEA/D used to identify the optimal learning rate and weight-decay values. In addition, Focal Loss was adopted to mitigate class-imbalance effects. The MOEA/D-based tuning of learning rate and weight decay substantially improved training efficiency and generalization, particularly under complex motif categories and heterogeneous backgrounds, enabling faster convergence and reducing overfitting.

In addition, to improve generalization during training, all layers except the final fully connected layer were kept frozen, and only the top-level parameters were fine-tuned. This strategy allows the model to extract features more stably, reduces the risk of overfitting, and improves recognition robustness.

Model parameters were optimized using the Adam optimizer, a widely adopted adaptive learning-rate method that effectively handles sparse gradients and adjusts learning rates for different parameters. The initial learning rate was set to 1×10^{-4} , together with weight decay (0.001) to mitigate overfitting. We further employed a ReduceLROnPlateau scheduler, which dynamically adjusts the learning rate based on the macro-F1 score on the validation set. Specifically, when the validation macro-F1 did not improve for five consecutive epochs, the scheduler reduced the learning rate by 50%, facilitating finer convergence in later training stages.

The dataset exhibits class imbalance, with relatively few samples in categories such as lion, deer, peacock, and tiger. This imbalance can degrade performance in multi-label classification, as the model may bias predictions toward majority classes, resulting in reduced accuracy and macro-F1. To alleviate this issue, Focal Loss was adopted as the training objective. This loss function explicitly addresses class imbalance by reweighting samples, enhancing the distinction between easy and hard examples so that the model focuses more on difficult-to-classify instances. The parameters α and γ in Focal Loss were computed and adjusted dynamically to better accommodate the uneven class distribution. Focal Loss is commonly used for imbalanced classification tasks, as it increases the contribution of hard samples while down-weighting those that are already well classified. The formulation of Focal Loss is given by:

$$FL(p_i) = -\alpha_i(1 - p_i)^\gamma \log(p_i) \tag{4}$$

In Eq. (4), p_i is the model-predicted probability for the true class; α_i is the class-balancing factor (typically set according to the number of samples per class); and γ is a focusing parameter that adjusts the relative weight between easy and hard examples.

Training was conducted for 100 epochs with a batch size of 64. After each epoch, model performance was evaluated on the validation set to monitor learning dynamics. With this adaptive learning-rate strategy, the

Table 2 | Final 100-epoch evaluation of the two hyperparameter sets identified by MOEA/D

Number	lr	wd	val_loss	hamming_acc	Macro f1
1	0.0005	0.000356737423401833	0.1759	93.79%	0.4761
2	0.000487687745512409	0.000717262109887259	0.1044	94.48%	0.6372

Table 3 | Runtime comparison before and after model optimization

Metric	ResNet50	Optimized ResNet50
Wall time	7206.8050	7222.6021
Avg epoch time	71.8370	71.9984

model maintains an appropriate learning pace at different training stages, improving efficiency while reducing overfitting.

To further enhance performance under multi-objective constraints, we optimized architectural hyperparameters using MOEA/D, aiming to maximize the F1 score and Hamming accuracy while minimizing the loss function. Although no fixed weights were explicitly assigned to individual objectives, the algorithm performed optimization by directly minimizing the negative values of the target functions: F1 and Hamming accuracy were converted to negative terms for maximization, whereas loss was treated as a minimization objective. To ensure optimization effectiveness, the maximum number of iterations was set to 20 generations. Solution diversity was regulated by tracking the number of non-dominated solutions; when this number exceeded 40, a subset of poorer solutions was automatically removed. Throughout the optimization process, stability and continuity were maintained by periodically saving states and enabling recovery from interruptions. Weight vectors for MOEA/D were generated using a random-direction method; specifically, the weight vector for each sub-problem is given in Eq. (5):

$$\lambda^{(i)} = \frac{u}{\|u\|_2}, u \sim U(0, 1)^{n_{Obj}} \tag{5}$$

That is, a random vector is sampled from $[0, 1]^{n_{Obj}}$ and normalized by its L_2 norm. This strategy generates randomly distributed subproblem directions in the objective space and serves the decomposition of multi-objective optimization. We further adopted the weighted Tchebycheff decomposition to scalarize multiple objectives:

$$g(x|\lambda) = \max_{1 \leq j \leq m} \lambda_j |f_j(x) - z_j^*| \tag{6}$$

where z^* denotes the current estimate of the ideal point. This decomposition can effectively handle non-convex Pareto fronts and provides distinct search directions for different subproblems. To enhance search diversity, Gaussian mutation was performed in logarithmic space. For each decision variable, mutation was applied with probability $p_{mut} = 0.20$, and the mutation perturbation followed $\Delta \sim N(0, \sigma^2)$, $\sigma = 0.25$. Afterwards, values were clipped in \log_{10} space to the predefined variable domain and then mapped back to the original space. This scheme is particularly suitable for scale-sensitive hyperparameters such as the learning rate.

MOEA/D first evaluated candidate hyperparameter ranges using 50 training epochs. Based on the dominance relations across the three objectives, two hyperparameter sets were selected as candidates. These two sets were then used to retrain the original model for 100 epochs each, and the three validation metrics for both sets were recorded (Table 2). Considering all three metrics jointly, with values retained to seven decimal places, the final configuration was determined as $lr = 4.877e - 4$ and $wd = 7.17e - 4$.

The results indicate that the learning rate (lr) exerts a markedly stronger influence on model performance than weight decay (wd). Across

all non-dominated solutions, the magnitude of lr variation shows a high correlation with improvements in Macro-F1. When lr is too low, the network struggles to escape underfitting; conversely, excessively high lr leads to a sharp rise in validation loss. Weight decay primarily affects generalization stability, functioning to reduce overfitting and to balance gradient contributions between majority and rare classes. The selected optimal configuration ($lr = 4.877e - 4$, $wd = 7.17e - 4$) therefore exhibits a pattern in which lr dominates performance gains while wd provides auxiliary support for generalization, consistent with typical deep-network behavior under small-sample conditions.

Model performance evaluation

To quantify the computational overhead introduced by the multi-objective optimization strategy, we compared training time between the baseline ResNet50 and the MOEA/D-optimized ResNet50 under identical training settings (same hardware, batch size, number of epochs, and data split). As shown in Table 3, the total wall-clock training time was 7206.8050 s for the baseline ResNet50 and 7222.6021 s for the optimized model. The average per-epoch training time was 71.8370 s and 71.9984 s, respectively. Differences in both total training time and per-epoch cost were below 0.3%, indicating that MOEA/D-based hyperparameter optimization occurs primarily during the pre-training search phase and imposes negligible additional computational burden on routine training of the final model. Thus, the proposed optimization achieves improved performance trade-offs while maintaining training efficiency comparable to the baseline.

To further validate the capability of the optimized model for recognizing Song-dynasty bronze-mirror motifs, we compared VGG16, EfficientNet-B0, the baseline ResNet50, and the MOEA/D-optimized ResNet50. Figure 6 reports their validation performance after 100 training epochs, with four panels corresponding to Macro-F1, Hamming accuracy, loss, and accuracy. As shown in Table 4 and Fig. 6, the optimized model exhibits clear advantages in accuracy. The classical VGG16 achieves Hamming accuracy close to that of ResNet50, but its loss remains substantially higher. The EfficientNet-B0 model, introduced in 2019, yields relatively low Hamming accuracy, although it outperforms VGG16 in terms of loss and Macro-F1. Overall, the optimized model shows more pronounced improvements in both accuracy and loss.

Further inspection of Fig. 6 indicates a clear train-validation mismatch for the comparative baselines. In Fig. 6c) EfficientNet-B0 and Fig. 6d) VGG16, validation accuracy increases only briefly at early epochs and then plateaus at a low level, while training accuracy continues to rise, revealing a pronounced generalization gap. This suggests that under a small, long-tailed, and highly imbalanced multi-label setting, some single-model baselines trained with a fixed strategy may converge to biased prediction regimes, limiting their validation-level discriminability. By contrast, the optimized ResNet50 shows more stable upward trends and convergence in validation Macro-F1 and Hamming accuracy, implying that multi-objective hyperparameter search helps mitigate “continued fitting on training but stagnation on validation,” thereby improving predictive consistency and robustness on held-out data.

To examine whether the performance gain of the optimized ResNet50 over the baseline ResNet50 is statistically significant, we conducted multiple independent repeated experiments for both models. All statistical analyses were performed in Python using open-source libraries including pandas and SciPy. Specifically, under a fixed train-validation-test split, we trained and evaluated each model with ten independent random seeds. For each run, Hamming accuracy and Macro-F1 on the validation set were recorded, from

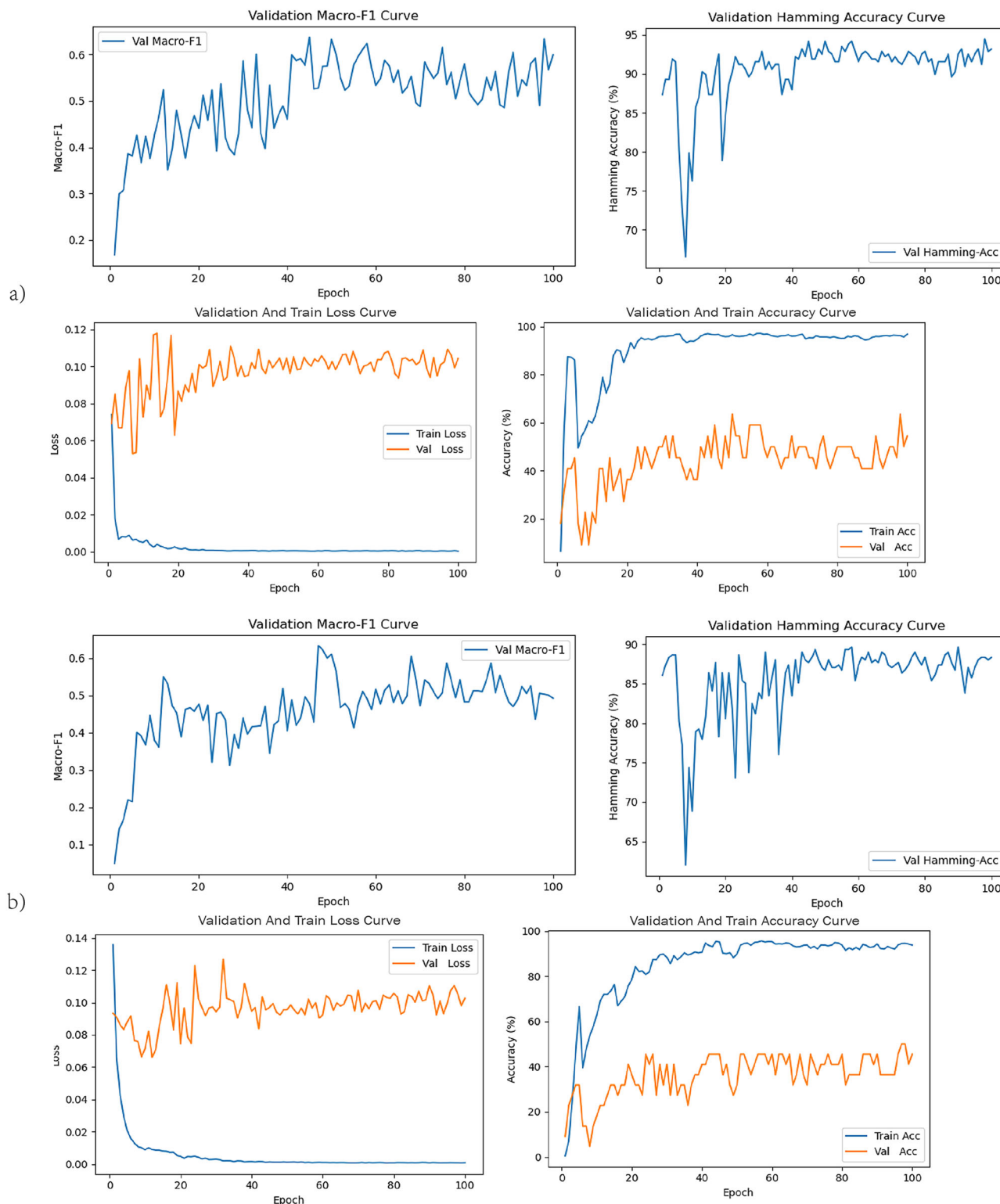


Fig. 6 | Evaluation metrics on the training and validation sets. The figures from top to bottom are (a) Optimized ResNet50; (b) ResNet50; (c) EfficientNet; (d) VGG16. The blue line represents the training set, and the orange line represents the validation set.

which sample means, standard deviations, and 95% confidence intervals were computed. We then applied a paired two-sided t-test to assess the significance of performance changes induced by the optimization strategy. Note that the random-seed repetitions were used to quantify sensitivity and stability with respect to training stochasticity under the same data split. Given the current dataset size and class imbalance, we did not perform k-fold cross-validation or external validation; therefore, generalization

across different splits remains to be tested on larger and more diverse datasets.

For Hamming accuracy, the optimized model achieves a significant improvement over the baseline (0.9383 vs. 0.8552, $p < 1e - 6$), indicating that the proposed optimized ResNet50 substantially enhances overall label-wise prediction consistency. By contrast, for Macro-F1, the mean score of the optimized model is slightly lower than that of the baseline (0.5322 vs.

Table 4 | Validation metrics of the four models after 100 training epochs

Modal	val_loss	Hamming accuracy	Macro f1
VGG16	0.7093	88.64%	0.1738
EfficientNet-B0	0.1138	66.23%	0.2586
ResNet50	0.1237	89.94%	0.4030
Optimized ResNet50	0.1044	94.48%	0.6372

0.5883, $p = 0.011$), with partially overlapping confidence intervals. We attribute this pattern primarily to class imbalance: Macro-F1 is highly sensitive to minority classes, whereas MOEA/D optimization oriented toward overall predictive accuracy tends to emphasize majority-class performance, which can induce mild fluctuations in rare classes. From an application standpoint, however, our emphasis lies on per-label accuracy and overall stability in practical deployment; thus, improvements in Hamming accuracy more directly reflect the real-world advantage of the optimized model. Taken together, the optimized model trades a modest decrease in Macro-F1 for a marked gain in label-wise reliability and training stability, improving its engineering suitability and practical value.

Moreover, the distribution of performance differences across repeated runs shows that gains from the optimized model are consistent across the primary metrics: the mean differences are positive, and the corresponding 95% confidence intervals do not cross zero. The paired significance tests ($p < 0.05$) indicate that the observed improvements are not attributable to random variation. These results therefore substantiate the effectiveness of the proposed optimization strategy, demonstrating that, with essentially unchanged training cost, it improves label-wise prediction quality and generalization robustness.

The core reason for the performance improvement of ResNet50 under MOEA/D lies in the algorithm's decomposition-based cooperative optimization, which enables more effective trade-offs among multiple objectives. Unlike single-objective or heuristic tuning, MOEA/D constructs a set of weighted subproblems so that different combinations of learning rate (lr) and weight decay (wd) are evaluated simultaneously against multiple performance criteria (Hamming accuracy, Macro-F1, and validation loss). During iteration, the neighborhood search mechanism preserves solution diversity and prevents premature convergence to local optima, while global crossover allows high-quality solutions to be shared across subproblems, thereby stably improving the efficiency of hyperparameter exploration.

In Fig. 7, the horizontal axis denotes the classes (Class 0–13) and the vertical axis denotes sample indices (22 samples in total). Green squares indicate labels predicted by the model, and pink polka dots indicate ground-truth labels; a pink polka dot overlapping a green square corresponds to a correct prediction. The optimized ResNet50 (labeled “modelresnetyh1” in the figure) shows a higher degree of overlap between predicted and true labels, indicating better sample-level performance. Nevertheless, misclassification persists in certain visually confusable categories, such as deer and peacock. Future work may consider incorporating transfer-learning strategies to further improve performance on these difficult minority classes.

In the multi-label recognition experiments on Song-dynasty bronze-mirror motifs, the optimized ResNet50 exhibits more stable overall performance and a more balanced capacity for class discrimination. On the same test set, the optimized ResNet50 achieves a macro-averaged F1 (macro-F1) of 0.358 and a Hamming accuracy of 0.919. These results outperform the baseline ResNet50 (macro-F1 0.347, Hamming accuracy 0.870), VGG16 (macro-F1 0.210, Hamming accuracy 0.903), and EfficientNet-B0 (macro-F1 0.260, Hamming accuracy 0.666), placing the optimized model in a leading position across the combined metrics. In Table 5, TP, FP, FN, and TN are four standard evaluation terms used to characterize the classification outcomes of different sample types during prediction; they are typically employed to derive precision, recall, and F1 score.

Taken together, Table 5 and Fig. 8 show that the optimized ResNet50 (denoted as “modelresnetyh1” in the figure) attains relatively high and

smoother F1 values across multiple classes. Relative to the baseline, the optimized model substantially reduces inter-class F1 volatility (the max–min gap decreases from 0.87 to 0.60), indicating that the optimization strategy alleviates the baseline pattern of “minority-class collapse and majority-class overfitting.” This yields more stable and robust recognition across motif categories. Concretely, the optimized model maintains high precision and recall for classes 0, 4, 6, 8, 11, and 13, producing F1 values largely within the 0.60–1.00 range (Table 5d), consistent with the corresponding high-intensity regions in Fig. 8. By contrast, although the baseline ResNet50 reaches $F1 \approx 1.00$ for a few classes (e.g., classes 3, 6, and 8), it shows cliff-like drops to $F1 = 0$ in classes 1, 2, 5, 7, 9, 10, and 12 (Table 5c), greatly amplifying inter-class performance disparity. This pattern suggests that the baseline model is more sensitive to the training-sample distribution and prone to systematic misses (excess FN) or complete failure in classes with scarce samples or weak motif cues. VGG16 attains a precision of 1.00 in some categories but exhibits low recall, reflecting conservative prediction behavior. EfficientNet-B0 achieves a recall close to or at 1.00 for many classes, yet its precision is markedly lower, with numerous false positives, leading to poor overall F1.

Across evaluation metrics, the optimized ResNet50 performs most strongly. In particular, it leads other models on F1 score, precision, and recall for most major categories, and shows superior precision–recall balance in more challenging classes (e.g., classes 6 and 13). Comparisons with other models indicate that the optimized ResNet50 has advantages in capturing fine-grained details and distinguishing motif categories, especially under complex pattern configurations; it more effectively balances precision and recall and avoids overfitting.

Regarding recognition errors, Table 5 and Fig. 8 jointly indicate two main sources of confusion. The first arises from visually similar animal motifs. As observed among neighboring classes or within the same “animal subtype,” these categories often show a combination of relatively high recall but impaired precision. Statistically, this corresponds to elevated FP (decreased precision) while recall remains at a moderate level, implying that the model can detect the higher-level semantic concept of “animal motif” but struggles with fine-grained differentiation. For bird-related motifs, for instance, different avian or mythic-bird patterns share locally similar contours, feather arrangements, and postural compositions. Such similarities lead to overlap in feature space and produce mutual misclassification between fine-grained categories (e.g., “bird” versus “phoenix”). This error type reflects blurred fine-grained semantic boundaries: the model is sensitive to coarse semantic cues but lacks sufficiently discriminative detail-level features.

The second source involves rare classes with very limited samples, leading to confusion or missed detections. Classes that persistently exhibit low F1 in Fig. 8 align closely with those showing low TP and high FN in Table 5, indicating that these errors are primarily driven by long-tail data scarcity. With small training sets and substantial intra-class variation, the model cannot learn stable discriminative representations for rare motifs. During inference, such classes are either attracted to visually similar high-frequency categories (manifested as increased FP and low precision) or not detected at all (increased FN and low recall), resulting in low or near-zero F1. Although the optimized model reduces the number of classes with extreme failure relative to the baseline, a small subset of long-tail categories still shows persistently low F1, underscoring sample scarcity as a central bottleneck for fine-grained recognition.

To mitigate training bias and uncertainty in generalization arising from limited samples in rare categories, we applied targeted generative data augmentation to a set of underrepresented classes (*Aix galericulata*, cattle, deer, lion, peacock, tiger, and turtle) and compared model performance before and after augmentation under an identical evaluation protocol. Given the constraints imposed by small datasets, we did not employ unconstrained generative models. Instead, we used a structure-preserving synthetic pipeline designed to remain consistent with basic physical plausibility. The pipeline applies set-level acquisition variation, localized appearance degradation attributable to wear or partial loss, and illumination changes. These



Fig. 7 | Prediction results on the test set. The horizontal axis denotes the classes (Class 0–13), and the vertical axis denotes sample indices (22 samples in total). Green squares indicate labels predicted by the model, and pink polka dots indicate ground-truth labels.

transformations emulate differences in the image formation process while preserving the geometric configuration of the motif, such that the resulting samples can be regarded as variants of the same motif under different imaging conditions. All synthesized samples were subsequently filtered

using mask-based SSIM and gradient-informed pattern constraints to reduce the likelihood of unrealistic artifacts and distributional shifts.

For evaluation, all images underwent the same preprocessing: resizing to 224 followed by center cropping to 224 × 224, and

Table 5 | Prediction metrics of the four models on the test set

All numerical values are reported to five decimal places							
a) VGG16							
Class	TP	FP	FN	TN	Precision	Recall	f1
Class 0	0	0	1	21	0	0	0
Class 1	0	2	2	18	0	0	0
Class 2	0	0	1	21	0	0	0
Class 3	0	0	1	21	0	0	0
Class 4	1	0	2	19	1	0.33333	0.5
Class 5	0	0	1	21	0	0	0
Class 6	5	6	0	11	0.45454	1	0.625
Class 7	1	0	2	19	1	0.33333	0.5
Class 8	1	0	0	21	1	1	1
Class 9	0	0	1	21	0	0	0
Class 10	0	0	1	21	0	0	0
Class 11	1	3	1	17	0.25	0.5	0.33333
Class 12	0	0	1	21	0	0	0
Class 13	1	0	3	18	1	0.25	0.4
b) EfficientNet-B0							
Class	TP	FP	FN	TN	Precision	Recall	f1
Class 0	0	2	1	19	0	0	0
Class 1	1	11	1	9	0.08333	0.5	0.14286
Class 2	0	4	1	17	0	0	0
Class 3	1	6	0	15	0.14286	1	0.25
Class 4	2	11	1	8	0.15385	0.6667	0.25
Class 5	0	1	1	20	0	0	0
Class 6	3	2	2	15	0.6	0.6	0.6
Class 7	0	2	3	17	0	0	0
Class 8	1	0	0	21	1	1	1
Class 9	1	20	0	1	0.04762	1	0.09091
Class 10	1	1	0	20	0.5	1	0.66667
Class 11	2	7	0	13	0.22222	1	0.36364
Class 12	1	20	0	1	0.04762	1	0.09091
Class 13	2	2	2	16	0.5	0.5	0.5
c) ResNet50							
Class	TP	FP	FN	TN	Precision	Recall	f1
Class 0	1	1	0	20	0.5	1	0.66667
Class 1	0	1	2	19	0	0	0
Class 2	0	1	1	20	0	0	0
Class 3	1	0	0	21	1	1	1
Class 4	3	9	0	10	0.25	1	0.4
Class 5	0	0	1	21	0	0	0
Class 6	4	0	1	17	1	0.8	0.88889
Class 7	0	0	3	19	0	0	0
Class 8	1	0	0	21	1	1	1
Class 9	0	0	1	21	0	0	0
Class 10	0	1	1	20	0	0	0
Class 11	2	3	0	17	0.4	1	0.57143
Class 12	0	0	1	21	0	0	0
Class 13	4	5	0	13	0.4444414	1	0.61539

Table 5 (continued) | Prediction metrics of the four models on the test set

d) Optimized ResNet50							
Class	TP	FP	FN	TN	Precision	Recall	f1
Class 0	1	0	0	21	1	1	1
Class 1	0	1	2	19	0	0	0
Class 2	0	2	1	19	0	0	0
Class 3	0	0	1	21	0	0	0
Class 4	3	4	0	15	0.42857	1	0.6
Class 5	0	0	1	21	0	0	0
Class 6	5	2	0	15	0.71429	1	0.83333
Class 7	1	1	2	18	0.5	0.33333	0.4
Class 8	1	0	0	21	1	1	1
Class 9	0	0	1	21	0	0	0
Class 10	0	0	1	21	0	0	0
Class 11	2	2	0	18	0.5	1	0.66667
Class 12	0	0	1	21	0	0	0
Class 13	4	1	0	17	0.8	1	0.88889

normalization using ImageNet statistics (mean = [0.485, 0.456, 0.406], std = [0.229, 0.224, 0.225]). The multi-label classifier produced per-class probabilities via a sigmoid transformation. We used class-wise PR-AUC (average precision; AP) as the primary threshold-independent metric, with F1 reported as a complementary measure. For rare classes, we additionally report grouped statistics relative to non-rare classes. To avoid tuning on the test set, F1 thresholds were selected using per-class thresholds determined on the validation set. In the test set, for true-positive labels belonging to the augmented categories, the post-augmentation model increased the mean predicted probability by +0.0187 (95% CI: +0.0082 to +0.0711; number of true-positive labels $N = 22$). To quantify changes in retrieval performance for rare classes, we defined the AP difference as $dAP = AP_{\text{before}} - AP_{\text{after}}$, such that negative dAP indicates improved AP after augmentation. Under this definition, the augmented categories showed a mean dAP of -0.0327 , whereas non-augmented categories showed a mean dAP of 0.0026 (difference: -0.0353), consistent with a positive shift in AP associated with targeted augmentation. Given the limited scale of the test set, we interpret these findings as supportive rather than definitive. Larger experiments are required to establish statistical stability and to more comprehensively characterize the effects of generative augmentation, in line with best practices emphasizing uncertainty-aware evaluation in artificial intelligence²⁸.

Building on this setting, we further replaced standard classification training with contrastive learning focused on motif regions, and added shortcut-detection and interpretability analyses, including background replacement, occlusion sensitivity, and the proportion of Grad-CAM attribution within the mask. In Table 6, metrics for Model A (the optimized model) and Model B (the optimized model additionally trained with synthetic images) show a consistent pattern: occluding the motif region (mask-pattern) results in a substantially larger decrease in AP than occluding the background (mask-background). This agreement suggests that the model's discriminative signals are primarily derived from the motif region, providing convergent evidence for the plausibility of the learned representations under extremely limited sample conditions.

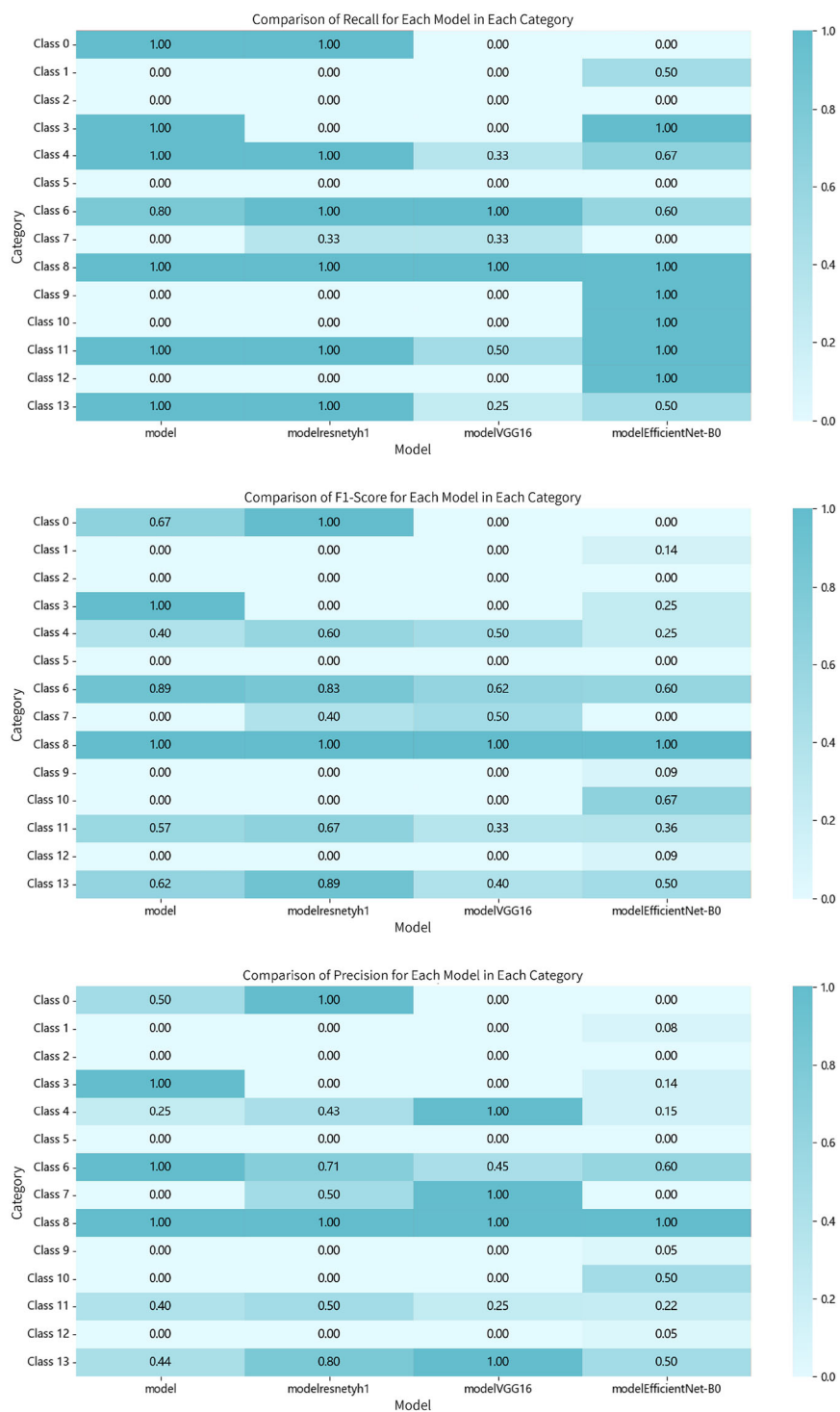


Fig. 8 | Heatmaps of three evaluation metrics across the 14 categories for the four models. The larger the value in the chart, the darker the blue color; the smaller the value, the lighter the blue color.

Table 6 | Occlusion-based evaluation of feature dependency (Models A and B) macroAP is computed over classes with at least two positive test samples (support ≥2)

Modal	Original macroAP	Mask-pattern macroAP	Mask-background macroAP
A	0.774	0.411	0.564
B	0.746	0.353	0.593

Discussion

Table 7 presents representative examples of 14 Song-dynasty bronze-mirror motifs, covering all 14 categories. Relative to the other models, the optimized ResNet50 achieves higher recognition accuracy and markedly reduces misclassifications. Across categories, the optimized ResNet50 more precisely identifies the motifs present in each image, highlighting its advantage in handling complex decorative patterns. By contrast, although EfficientNet-B0 predicts multiple categories for nearly every image, its

Table 7 | Image-recognition results of the four models















Number	Image	True label	VGG16	EfficientNet-B0	ResNet50	Optimized ResNet50
1		<i>Aix galericulata</i>		Lion Tiger	<i>Aix galericulata</i>	<i>Aix galericulata</i>
2		Bird		Crane Lion Phoenix Tiger	Phoenix	Phoenix
3		Butterfly Dragon Phoenix	Dragon Phoenix	Bird Cattle Crane Dragon Lion Phoenix Tiger	Phoenix	Phoenix
4		Cattle	Dragon	Cattle Lion Tiger		Dragon
5		Deer	dragon	Bird Crane Dragon Lion tiger	Crane Turtle	
6		Dragon	Dragon	Dragon Lion tiger	Dragon	Dragon
7		Dragon	Dragon Phoenix	<i>Aix galericulata</i> Bird Cattle Crane Dragon Lion Phoenix Tiger		Dragon
8		Fish	Fish			Fish
9		Goose	Goose	Butterfly Goose Lion Tiger	Goose	Goose
10		Lion		Crane Lion Tiger		
11		Peacock	Dragon	<i>Aix galericulata</i> Bird Butterfly Cattle Crane Deer Lion Peacock Phoenix Tiger		Crane
12		Phoenix	Dragon	Bird Butterfly Cattle Crane Fish Lion Phoenix Tiger Turtle	Phoenix	Phoenix

Table 7 (continued) | Image-recognition results of the four models

Number	Image	True label	VGG16	EfficientNet-B0	ResNet50	Optimized ResNet50
13		Tiger		Bird Crane Fish Lion Peacock Tiger		
14		Crane Turtle	Bird	Bird Lion Tiger	Crane Turtle	Crane Turtle

accuracy is lower, and misclassifications are more frequent, indicating a higher false-positive rate.

These results indicate that the optimized ResNet50 offers stronger generalization and robustness under fine-grained, inter-class imbalanced conditions for Song-dynasty bronze-mirror motif recognition, making it a suitable primary backbone model or core component within an ensemble system for this task.

We propose a framework that integrates a multi-objective evolutionary algorithm (MOEA/D) with a deep convolutional neural network (ResNet50) for automated recognition of animal motifs on Song-dynasty bronze mirrors. Through hyperparameter optimization, the resulting model improves overall predictive consistency and Hamming accuracy. Although the minority-class-sensitive Macro-F1 exhibits minor fluctuations—primarily attributable to class imbalance—the optimized model shows enhanced stability and generalization. Compared with traditional handcrafted feature-extraction approaches, the deep-learning framework automatically learns high-level motif representations, reducing errors introduced by human intervention and lowering labor demands. Despite the favorable experimental results, dataset size and class imbalance remain key constraints on further improvement. Future work may enhance generalization by increasing samples for minority classes, adopting transfer-learning strategies, or using generative models to synthesize additional training data. Cross-split validation on larger, more diverse collections of bronze-mirror motifs will also be required to establish robustness more conclusively. Moreover, given the computational cost associated with optimization, improving algorithmic efficiency represents another direction for continued study. Overall, this work provides an efficient solution for motif recognition in cultural-heritage research, with particular relevance to automated identification and digital preservation of Song-dynasty bronze mirrors. With expanded datasets and further advances in optimization methods, the proposed approach may be extended to other heritage domains to support intelligent conservation and analysis. At the same time, high recognition accuracy alone is insufficient in cultural-heritage contexts; models should also provide interpretations consistent with historical backgrounds and aesthetic values. Accordingly, future research will prioritize explainable AI techniques, such as Grad-CAM and LRP-based visualization, to uncover the key features driving model decisions.

Data availability

The datasets generated and/or analyzed during the current study are not publicly available due to ongoing in-depth, multi-perspective investigations by the research team, but are available from the corresponding author on reasonable request.

Code availability

The underlying code used and/or analyzed in the current study is available from the corresponding author upon reasonable request.

Received: 26 September 2025; Accepted: 27 February 2026;

Published online: 16 March 2026

References

- Jiang, S. *Tongjing* (Huangshan Shushe, 1995).
- Guo, Y. The aesthetic of brightness in Han mirror inscriptions. *J. Am. Orient. Soc.* **141**, 93–124 (2021).
- Zhang, M. Auspicious animals in bronze mirrors of Sui and Tang dynasties. *Artseduca* **40**, 12–30 (2024).
- Cammann, S. The lion and grape patterns on Chinese bronze mirrors. *Artibus Asiae* **16**, 265–291 (1953).
- Varsano, P. Disappearing objects/elusive subjects. *Representations* **124**, 96–124 (2013).
- He, Q., Zhao, Q. & Wu, Z. Image processing and restoration of cultural relics of Song dynasty in Sichuan based on 3D modeling technology. In *International Conference on Cognitive based Information Processing and Applications* (CIPA 2021) (eds J. Jansen, B., Liang, H. & Ye, J.) 115–121 (Springer, 2022).
- Bulling, A. G. & Drew, I. The dating of Chinese bronze mirrors. *Archives of Asian Art* **25**, 36–57 (1972).
- Wu, K. Viewing the concept of creation and genetic mapping of the traditional weaving process from the perspective of utensils—taking the flat barrel cage as an example. *Adv. Design Res.* **1**, 87–93 (2023).
- Qi, X., He, X., Chen, S. W. & Hai, T. A framework of evolutionary optimized convolutional neural network for classification of Shang and Zhou dynasties bronze decorative patterns. *PLoS ONE* **19**, e0293517 (2024).
- Pathak, A. R., Pandey, M. & Rautaray, S. Application of deep learning for object detection. *Procedia Comput. Sci.* **132**, 1706–1717 (2018).
- Xiao, Y. et al. A review of object detection based on deep learning. *Multimed. Tools Appl.* **79**, 23729–23791 (2020).
- Li, Y. et al. Detection and recognition of Chinese porcelain inlay images of traditional Lingnan architectural decoration based on YOLOv4 technology. *Herit. Sci.* **12**, 1–41 (2024).
- Wu, M., Yang, L. & Chai, R. Research on multi-scale fusion method for ancient bronze ware X-ray images in NSST domain. *Appl. Sci.* **14**, 4166 (2024).
- Yohannes, Rivan, M. E. A., Devella, S. & Tinaliah. A novel optimization strategy for CNN models in Palembang Songket motif recognition. *Int. J. Adv. Comput. Sci. Appl.* **16**, 830 (2025).
- Xu, L., Lu, L. & Liu, M. Construction and application of a knowledge graph-based question answering system for Nanjing Yunjin digital resources. *Herit. Sci.* **11**, 1–17 (2023).
- Fu, Y., Shi, K. & Xi, L. Artificial intelligence and machine learning in the preservation and innovation of intangible cultural heritage: ethical considerations and design frameworks. *Digit. Scholarsh. Humanit.* **40**, 487–508 (2025).
- Li, M., Zhou, G. & Li, Z. Fast recognition system for tree images based on dual-task Gabor convolutional neural network. *Multim. Tools Appl.* **81**, 28607–28631 (2022).
- Liu, Y., Pang, C., Zhan, Z., Zhang, X. & Yang, X. Building change detection for remote sensing images using a dual-task constrained deep siamese convolutional network model. *IEEE Geosci. Remote Sens. Lett.* **18**, 811–815 (2021).

19. Wu, F. et al. Wavelet-based dual-task network. *IEEE Trans. Neural Netw. Learn. Syst.* **36**, 1–14, <https://doi.org/10.1109/TNNLS.2024.3486330> (2024).
20. Cai, P., Zhang, Y., He, H., Lei, Z. & Gao, S. DFNet: a differential feature-incorporated residual network for image recognition. *J. Bionic Eng.* **22**, 931–944 (2025).
21. Zhang, X. & Ding, T. Style classification of media painting images by integrating ResNet and attention mechanism. *Heliyon* **10**, e27178 (2024).
22. Shakibhamedan, S., Amirafshar, N., Baroughi, A. S., Shahhoseini, H. S. & TaheriNejad, N. A. C. E.-C. N. N. Approximate carry disregard multipliers for energy-efficient CNN-based image classification. *IEEE Trans. Circuits Syst. I Regul. Pap.* **71**, 2280–2293 (2024).
23. Manakitsa, N., Maraslidis, G. S., Moysis, L. & Fragulis, G. F. A review of machine learning and deep learning for object detection, semantic segmentation, and human action recognition in machine and robotic vision. *Technologies* **12**, 15 (2024).
24. Qin, J., Pan, W., Xiang, X., Tan, Y. & Hou, G. A biological image classification method based on improved CNN. *Ecol. Inform.* **58**, 101093 (2020).
25. Zhang, Q. & Li, H. MOEA/D: a multiobjective evolutionary algorithm based on decomposition. *Trans. Evol. Comput.* **11**, 712–731 (2007).
26. Luo, J., Jiao, L., Shang, R. & Liu, F. Learning simultaneous adaptive clustering and classification via MOEA. *Pattern Recognit.* **60**, 37–50 (2016).
27. Wang, Q., Gu, Q., Chen, L., Guo, Y. & Xiong, N. A MOEA/D with global and local cooperative optimization for complicated bi-objective optimization problems. *Appl. Soft Comput.* **137**, 110162 (2023).
28. Lomas, J. D. et al. Evaluating the alignment of AI with human emotions. *Adv. Design Res.* **2**, 88–97 (2024).

Acknowledgements

This research was supported by the National Social Science Foundation of China (Arts Project) (24EG247).

Author contributions

Q.F. was responsible for the overall direction of the research and writing and provided guidance. K.Y. was responsible for writing the manuscript and

conducting the main research. Y.L. was responsible for data collection and assisted with the research. T.M. was responsible for the technical control of the model setup and the compilation of experimental results. All listed authors made substantial, direct, and intellectual contributions to the work and approved its publication.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to Kexin Yu.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2026