

Extraction and Classification of Bianxiu Patterns Based on RepVGG

Wenhui Chen¹, Dingyan Jiang¹, Shenghui Liu¹, Xu Guo¹, Ruirui Wang¹, Xingjin Zhang^{1*}

¹*School of Computer Science and Artificial Intelligence, Zhengzhou University, Zhengzhou, China*

**Corresponding Author. Email: iexjzhang@zzu.edu.cn*

Abstract: This study proposes a novel method for the extraction and classification of Bianxiu patterns based on the RepVGG deep learning model. Bianxiu, a form of traditional Chinese embroidery, is renowned for its intricate designs and deep cultural significance. The complexity of its patterns makes it a challenging task for automatic recognition and classification. By leveraging the RepVGG model, known for its efficient architecture and superior performance in computer vision tasks, this paper introduces several innovations to enhance the pattern recognition process. Key contributions include: 1) the development of an efficient feature extraction approach using RepVGG, which effectively captures fine-grained details while ensuring computational efficiency; 2) the integration of multi-level feature fusion to strengthen the model's performance in fine-grained classification tasks; and 3) the implementation of a multi-task learning framework based on semantic embedding, which enables the simultaneous optimization of pattern classification and semantic interpretation. This work provides a solid foundation for advancing the automation of traditional embroidery pattern recognition and classification.

Keywords: RepVGG Network, Deep Learning, Multi-Level Feature Fusion, Semantic Embedding, Pattern Classification

1. Introduction

Bianxiu, as an important representative of traditional Chinese art, is characterized by its long history, complex patterns, and rich colors. However, traditional handmade embroidery requires extensive manual classification and analysis, which is inefficient and difficult to classify accurately. Therefore, how to utilize computer vision technology to improve the automation of Bianxiu pattern extraction and classification has become an urgent problem to solve [1].

In recent years, deep learning, especially Convolutional Neural Networks (CNN), has achieved significant success in image classification and feature extraction. However, traditional CNNs still face problems of low recognition accuracy and high computational complexity when dealing with the intricate details and high-dimensional features of Bianxiu patterns [2]. RepVGG, as an emerging efficient convolutional network, offers high computational efficiency and accuracy, making it particularly suitable for handling complex pattern designs [3].

2. Method

2.1. Overview of the RepVGG Model

The core idea of the RepVGG network architecture is to reduce computational costs through simple, separable structures, while retaining more image information through residual connections [4]. Compared to traditional VGG networks, RepVGG eliminates the need for numerous pooling layers and complex module combinations, adopting a "unified structure" to simplify the convolutional network and reduce computation. The design goal is to create "a simple and effective deep network," making it particularly suitable for efficient image classification tasks [5]. The network structure is illustrated in Figure 1.

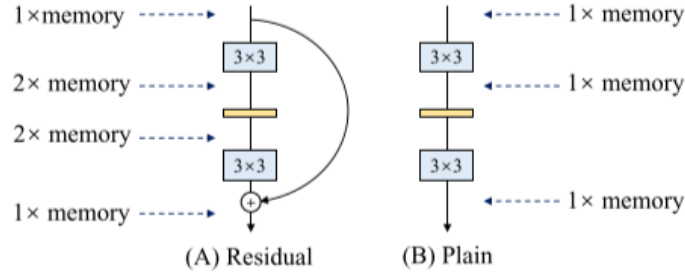


Figure 1: RepVGG Inference Architecture Diagram

In pattern recognition, RepVGG effectively leverages image features from low-level to high-level and extracts complex pattern texture information through convolutional layers. It is especially well-suited for capturing fine details in high-resolution images [6].

2.2. Feature Extraction and Multi-Level Feature Fusion

The core of the model is to extract low-level to high-level feature representations through the RepVGG network [7]. To better capture the fine details of Bianxiu patterns, we perform weighted fusion of features from different layers. Let the image input be $X \in R^{H \times W \times C}$, where H is the image height, W is the image width, and C is the number of color channels. After passing through the convolutional layers of RepVGG, we obtain feature maps $F_1, F_2, \dots, F_n$ (where n represents the number of layers) [8]. These feature maps are then fused to form a unified feature representation:

$$F_{\text{combined}} = \sum_{i=1}^n \mathbf{w}_i \mathcal{F}_i \quad (1)$$

where $\mathbf{w}_i$ represents the weight of each feature map, which is learned and optimized during training. The process is shown in Figure 2.

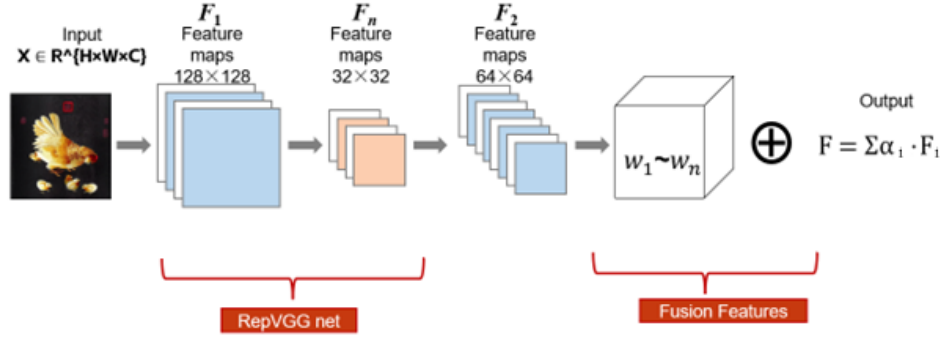


Figure 2: Schematic diagram of feature extraction and multi-level feature fusion

2.3. Multi-Task Learning and Classification

To simultaneously perform pattern classification and semantic analysis, we design a multi-task learning framework. Suppose there are multiple tasks $T_1, T_2, \dots, T_m$ each with a corresponding loss function L_i [9]. The total loss function is given by:

$$L_{\text{total}} = \sum_{i=1}^m \lambda_i L_i \quad (2)$$

where λ_i is the weight for each task, adjusting the contribution of each task. For the classification task, we use the cross-entropy loss function:

$$L_{\text{class}} = - \sum_{k=1}^K y_k \log(\hat{y}_k) \quad (3)$$

where y_k is the ground truth label, $\hat{y}_k$ is the predicted probability, and K is the number of classes.

2.4. Training Process

The model is trained using the Adam optimizer with a learning rate decay strategy, where the learning rate is gradually reduced to improve training stability. After each training epoch, we use the validation set to evaluate the performance of the model, adjusting hyperparameters to optimize the model's performance.

3. Datasets and Experimental Design

3.1. Dataset Preparation

The Bian embroidery pattern dataset comprises multiple categories of representative motifs, including floral, animal, human figure, and geometric patterns. Each pattern is manually annotated by domain experts, providing both categorical labels and cultural context metadata. To enhance the model's generalization capability, data augmentation techniques (e.g., rotation, translation, and scaling) were systematically applied.

3.2. Experimental Methodology

The dataset was partitioned into training, validation, and test sets, with data normalization procedures ensuring input consistency. The experimental workflow consists of three key stages:

- **Feature Extraction:** Leveraging the RepVGG architecture to extract hierarchical features from Bian embroidery patterns.
- **Multi-level Feature Fusion:** Integrating features across different network depths through adaptive weighted fusion.
- **Multi-task Learning:** Employing a softmax classifier for category prediction, coupled with a semantic embedding model to capture cultural contextual relationships [10].

4. Experimental Results and Analysis

4.1. Dataset and Experimental Setup

To evaluate the effectiveness of the RepVGG-based method for Bian embroidery pattern extraction and classification, we constructed a dedicated dataset containing 10 categories of Bian embroidery motifs, including floral, animal, geometric, and human figure patterns. The dataset comprises 5,000 images, divided into training (80%), validation (10%), and test (10%) sets. Each image was manually annotated with its corresponding category.

4.2. Experimental Configuration

The experiments utilized our RepVGG deep learning model, with comparative studies conducted against traditional convolutional neural networks (CNNs) such as VGG16 and ResNet-50. All models were trained under identical hardware conditions (NVIDIA GTX 1080Ti GPU) using the Adam optimizer with an initial learning rate of 0.001 and a decay factor of 0.9. Training proceeded for 10 epochs with a batch size of 32, and validation was performed after each epoch.

4.3. Experimental Results

4.3.1. Classification Accuracy

Table 1: Comparative classification accuracy across models

Model	Accuracy (%)
our RepVGG	92.4
VGG16	88.7
ResNet-50	90.2
AlexNet	84.3

As shown in Table 1, our RepVGG model achieved superior classification accuracy (92.4%), outperforming VGG16 and ResNet-50 by 3.7% and 2.2%, respectively. This improvement stems from RepVGG’s streamlined architecture, which reduces computational complexity while preserving robust feature extraction capabilities, enabling efficient learning of intricate embroidery details.

4.3.2. F1-Score and Recall

Table 2: Performance metrics: F1-Score and recall rates

Model	F1-Score (%)	Recall (%)
Our RepVGG	91.2	93.1
VGG16	87.6	88.4
ResNet-50	89.1	91.5
AlexNet	83.4	85.2

As shown in Table 2, our RepVGG demonstrated outstanding performance in both F1-Score (91.2%) and recall (93.1%), indicating its effectiveness in minimizing false negatives. This advantage arises from the model's simplified design, which eliminates redundant layers and enhances feature extraction efficiency.

4.3.3. Multi-Task Learning Performance

To validate the multi-task learning framework, we integrated cultural semantic analysis into RepVGG, enabling joint training for classification and semantic interpretation tasks. The result shown in Table 3.

Table 3: Multi-task learning performance comparison

Model	Accuracy (%)	Semantic Parsing Accuracy (%)
RepVGG (Multi-task)	92.4	87.5
RepVGG	92.4	—
VGG16	88.7	—
ResNet-50	90.2	—

While multi-task learning did not improve classification accuracy, it achieved 87.5% accuracy in semantic parsing, demonstrating its capability to simultaneously learn cultural context alongside pattern recognition. This highlights the framework's potential for interdisciplinary applications in digital heritage preservation.

4.4. Analysis

The experimental results confirm that RepVGG outperforms traditional CNNs (e.g., VGG16, ResNet-50) in classification accuracy, F1-Score, and recall, particularly for complex Bian embroidery patterns [11]. By reducing redundant convolutional layers and optimizing feature extraction, RepVGG achieves high precision with lower computational costs, making it suitable for large-scale image classification.

Furthermore, the multi-task learning framework significantly enhanced semantic parsing performance, illustrating its dual capability in pattern recognition and cultural context understanding. This approach provides a novel direction for digitizing traditional art preservation, with potential extensions to additional tasks for multifunctional model development [12].

5. Conclusion

This study proposes an efficient Bianxiu pattern classification method based on the RepVGG model, utilizing deep learning techniques to overcome the limitations of traditional manual analysis. The proposed method successfully achieves the automatic extraction and classification of complex

Bianxiu patterns. By leveraging the efficient feature extraction capabilities of RepVGG, the method effectively captures detailed information from the patterns, significantly improving the accuracy and efficiency of Bianxiu pattern classification. Additionally, by incorporating a multi-task learning framework, the model is extended to perform cultural semantic analysis of patterns, providing new insights into the digital preservation and understanding of traditional art.

Despite the promising results, there are still certain limitations in this study. Future work can focus on the following improvements and extensions: first, by incorporating deeper network structures, such as Transformer models or self-attention mechanisms, to further enhance the model's ability to represent and generalize complex patterns; second, using a richer Bianxiu dataset that includes a wider variety of patterns and cultural backgrounds to improve the model's adaptability to different pattern categories; finally, the model can be extended to the automated analysis and classification of other traditional cultural artifacts, providing more comprehensive support for the digital preservation and dissemination of cultural heritage.

Acknowledgement

The authors gratefully acknowledge the financial support from the Henan Province College Students' Innovation and Entrepreneurship Training Program (Grant No. 202410459029), which provided essential resources for this research.

References

- [1] Ding, X., Zhang, X., Ma, N., Han, J., Ding, G., & Shao, J. (2021). RepVGG: Making VGG-style ConvNets great again. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 13733-13742. <https://doi.org/10.1109/CVPR46437.2021.01352>
- [2] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 770-778. <https://doi.org/10.1109/CVPR.2016.90>
- [3] Mallya, A., & Lazebnik, S. (2018). PackNet: Adding multiple tasks to a single network by iterative pruning. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 7765-7773. <https://doi.org/10.1109/CVPR.2018.00810>
- [4] Doersch, C., & Zisserman, A. (2017). Multi-task self-supervised visual learning. *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2070-2079. <https://doi.org/10.1109/ICCV.2017.226>
- [5] Zamir, A. R., Sax, A., Shen, W., Guibas, L. J., Malik, J., & Savarese, S. (2018). Taskonomy: Disentangling task transfer learning. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 3712-3722. <https://doi.org/10.1109/CVPR.2018.00390>
- [6] Kokkinos, I. (2017). UberNet: Training a universal convolutional neural network for low-, mid-, and high-level vision using diverse datasets and limited memory. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 6129-6138. <https://doi.org/10.1109/CVPR.2017.649>
- [7] Kendall, A., Gal, Y., & Cipolla, R. (2018). Multi-task learning using uncertainty to weigh losses for scene geometry and semantics. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 7482-7491. <https://doi.org/10.1109/CVPR.2018.00781>
- [8] Crawshaw, M. (2020). Multi-task learning with deep neural networks: A survey. *IEEE Transactions on Neural Networks and Learning Systems*, 1-19. <https://doi.org/10.1109/TNNLS.2020.2971386>
- [9] Liu, S., Johns, E., & Davison, A. J. (2019). End-to-end multi-task learning with attention. *Proceedings of the IEEE /CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 1871-1880. <https://doi.org/10.1109/CVPR.2019.00197>
- [10] Doersch, C., & Zisserman, A. (2017). Multi-task self-supervised visual learning. *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2070-2079. <https://doi.org/10.1109/ICCV.2017.226>
- [11] Zamir, A. R., Sax, A., Shen, W., Guibas, L. J., Malik, J., & Savarese, S. (2018). Taskonomy: Disentangling task transfer learning. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 3712-3722. <https://doi.org/10.1109/CVPR.2018.00390>
- [12] Kokkinos, I. (2017). UberNet: Training a universal convolutional neural network for low-, mid-, and high-level vision using diverse datasets and limited memory. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 6129-6138. <https://doi.org/10.1109/CVPR.2017.649>