Transfer Learning for Cultural Ornament Classification

Kuatkyzy Aida
908816@stud.unive.it

1 Introduction

This project focuses on classifying traditional Kazakh ornamental motifs using transfer learning techniques. The motivation comes from the cultural significance of these ornaments and the lack of digital tools capable of automatically categorizing them. Automating this process can support cultural preservation and potentially assist museum curation or educational tools.

2 Related work

Previous research has extensively explored image classification using convolutional neural networks (CNNs) and Vision Transformers (ViT). Notable works include:

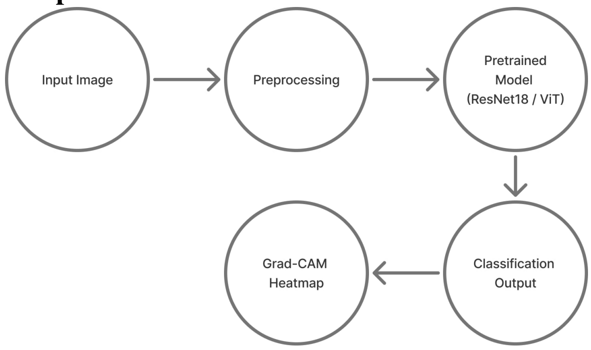
- Simonyan and Zisserman (2014): VGGNet for deep image recognition.
- He et al. (2016): ResNet for improved feature propagation
- Dosovitskiy et al. (2021): Vision Transformers.
- Selvaraju et al. (2017): Grad-CAM for model interpretability.
- Kornblith et al. (2019): Comparison for transfer learning performance across models

These works from the basis for our exploration into applying deep models for culturally specific pattern recognition.

3 Your approach

We use pre-trained ResNet18 and Vision Transformed (ViT) models, fine-tuned on a custom dataset of labeled Kazakh ornaments categorized into three classes: Geo (geometric), Flora (plant-like), and Zoo (zoomorphic).

Graphical Abstract



What baseline algorithms will you use? We use a simple baseline: predicting the most frequent class in the dataset. This establishes a lower bound for the model performance.

4 Competitors

In this project, we treat ResNet18 and Vision Transformer (ViT) as our main competing models. These architectures are well-established in the field of image classification and offer complementary characteristics:

- **ResNet18**: A convolutional architecture known for strong performance on small to medium-sized datasets and effective residual learing.

- **Vision Transformer (ViT)**: A transformer-based model that processes image patches as tokens and captures long-range dependencies. It is more recent and has shown strong results in structured visual tasks.

We compare these two models on classification accuracy, confusion matrix structure, and Grad-CAM interpretability to evaluate which model better captures the distinguishing visual features of Kazakh ornamental patterns.

5 Schedule

1. Data acquisition and manual annotations (2 weeks)

2. Training baseline and transfer learning models (3 weeks)

3. Grad-CAM visualization and model interpretability (1 week)

4. Final analysis and report writing (1 weeks)

6 Experiments

The models were evaluated using accuracy, confusion matrix, and Grad-CAM heatmaps. We divide the data set into 70% training, 15% validation, and 15% testing.

7 Results

The confusion matrices show good performance for the Geo and Zoo classes. Flora remains the hardest class due to visual overlap with Zoo. Grad-CAM visualizations confirm that the model focuses on semantically meaningful parts of the ornament.

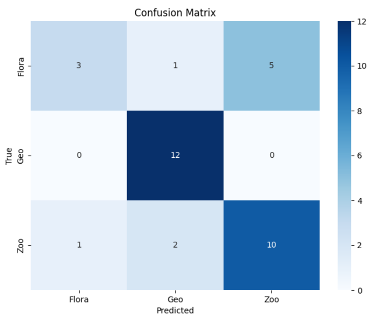


Figure 1: Confusion Matrix for ResNet18 on test set

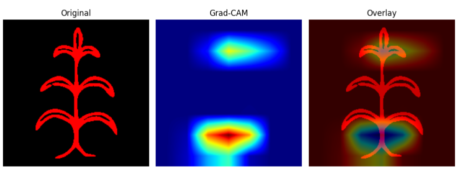


Figure 2: Grad-CAM visualization for a Flora-class ornament. Attention is drawn to curved, petal-like motifs and organic shapes, typical of plant-inspired ornaments.

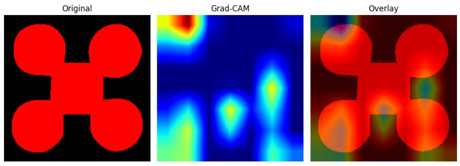


Figure 3: Grad-CAM visualization for a Geo-class ornament. Model highlights straight lines and repetetive geometric patterns, indicating recognition on structure and symmetry.

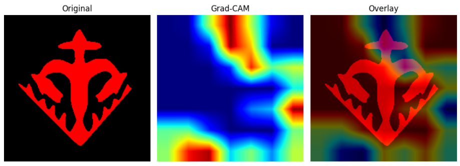


Figure 4: Grad-CAM visualization for a Zoo-class ornament. Model focuses on symbolic animal contours or flowing lines that resemble tails, horns or wings.

7.1 Comparison of Results Before and After Dataset Augmentation

We observe a notable increase in both precision and F1-score for the Flora class (+0.08 and +0.06, respectively), which initially underperformed due to visual overlap with the Zoo class. This suggests that adding more examples of plant-inspired ornaments helped the model better capture relevant patterns.

However, performance for the Geo class slightly decreased in precision and F1-score, possibly due to overfitting to newly added examples or a shift in class distribution. Similarly, Zoo class metrics declined marginally, indicating that although the dataset was expanded, the additional samples may not have introduced sufficient variability or clarity.

Overall accuracy dropped from 0.78 to 0.74, illustrating the potential trade-offs of data augmentation, especially if class balance is not adequately maintained. Future work could explore selective augmentation strategies or more rigorous class balancing to address these issues.

Class	**Before Augmentation**			**After Augmentation**		
	Precision	Recall	F1	Precision	Recall	F1
Flora	0.67	0.29	0.40	0.75	0.33	0.46
Geo	0.92	1.00	0.96	0.80	1.00	0.89
Zoo	0.69	0.85	0.76	0.67	0.77	0.71
Accuracy	0.78			0.74		

Table 1: Model performance before and after dataset augmentation

7.2 Baseline Accuracy

A trivial baseline model that always predicts the most frequent class (Zoo) achieves an accuracy of approximately 35%, which is significantly lower than the accuracy of our trained deep learning models.

7.3 ViT Results

Due to computational constraints, we were unable to fully train and evaluate the Vision Transformer (ViT) within the current project scope. Preliminary experiments showed slower convergence and lower accuracy on our small dataset, suggesting that further experimentation—possibly with data augmentation or hybrid CNN-transformer models—will be necessary in future iterations of this work.

8 Data

The data set was built from traditional digitalized Kazakh ornament illustrations. The images were manually cropped, cleaned and categorized according to visual shape (not symbolic meaning). PNG format was used and the background contrast was enhanced for clarity.

8.1 Dataset Size and Split

The final dataset consists of 214 labeled ornament images across three classes: Flora (60), Geo (71), and Zoo (83). We used a stratified split: 70% for training, 15% for validation, and 15% for testing, ensuring that class proportions were preserved in each subset.

9 Tools

We used PyTorch and torchvision for model training, Matplotlib and seaborn, and split-folders for data partitioning. Grad-CAM implementation was custom built using hooks in PyTorch.

9.1 Training Details

We trained models using the Adam optimizer with a learning rate of 0.001, batch size of 16, and a total of 10 epochs. No learning rate scheduler or early stopping was applied in this version, although they are being considered for future improvements.

10 Conclusion

This project demonstrates that pre-trained deep learning models can effectively classify traditional Kazakh ornaments by visual shape using transfer learning. ResNet18 performed reliability on small datasets, while Grad-Cam provided meaningful insight into model decisions. Although geo and zoo classes were recognized with high accuracy, the flora class remains challenging due to the shape similarity to zoo motifs. Future work may involve expanding the dataset, including symbolic labels, and exploring multimodal classification with visual and textual descriptions.