Deep Metric Learning for Ancient Coin Identification

JMU Department of Computer Science
JMU Department of Engineering

By: Dhanshree Atre, Trevor Schonbrun

Advisors: Dr. Nathan Sprague and Dr. Jason Forsyth

Acknowledgements: Jackson Greer, Quinnie Lu

Department of Computer Science, James Madison University, Harrisonburg, Virginia 22807

JAMES MADISON UNIVERSITY 1908

Introduction

Establishing provenance is essential for verifying the authenticity, legality, and value of ancient coins. Manually searching auction catalogs is slow, error-prone, and requires expert knowledge. We address this problem using deep metric learning to match coins with catalog images. Our task demands training on many coins with multiple photos of each, so the model learns to recognize the same coin across varying lighting, angles, and wear. Since catalogs typically provide only one image per coin, data augmentation is critical for generating the diversity needed for consistent recognition.

Figure 1 – Which coins(a) on the right matches the coin on the left?

Motivation

- JMU's Sawhill Collection holds over 400 ancient Greek and Roman coins, donated to the university following Dr. John Sawhill's death in 1976
- Much of the original 1,300-coin collection was sold at auction in 1979, leaving many provenance records incomplete or lost
- Incomplete records are a widespread problem faced by museums and collectors worldwide
- Automating coin-to-catalog matching could help recover lost provenance at scale

Metric Learning vs. Classification

Traditional classification focuses on predicting which predetermined class an instance belongs to. The downside is that it can only separate training classes Metric learning instead learns an embedding space where images of the same coin are pulled close together while images of different coins are pushed apart, making distance a direct measure of similarity.

In our case, we want to know if two images correspond to the same physical coin, not if they are the same type, so metric learning is a better fit. Metric learning is also common in face recognition, which has similar goals and challenges:

- Verification - want to know if two images are the same person/coin
- Open set recognition
- Variability between images of the same identity
- Similarities between different people/coins

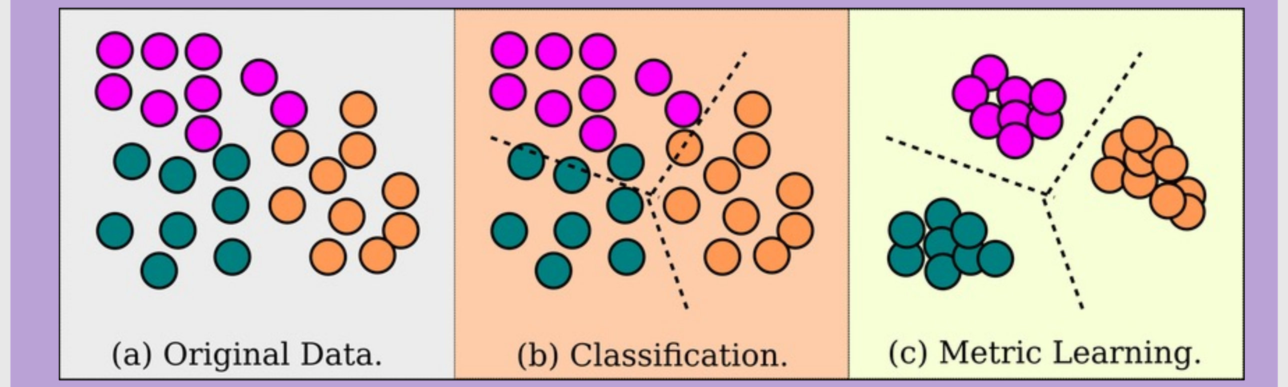

Figure 2 – Classification vs Metric learning [1]

Data Augmentation

Since each coin has only one image, we need to artificially increase the number of images by performing various data augmentations. The augmentation we perform are a random combination of:

- Brightness
- Contrast
- Saturation
- Gaussian blur
- Rotation
- Translation
- Bilinear interpolation
- Perspective distortion

Figure 3 – Example augmentations

Architecture Truncation

ResNet50's final layers aggregate spatial information into a single global vector, discarding location. Since our coin images are spatially aligned, we hypothesized that removing those layers would preserve more useful spatial structure and improve matching accuracy.

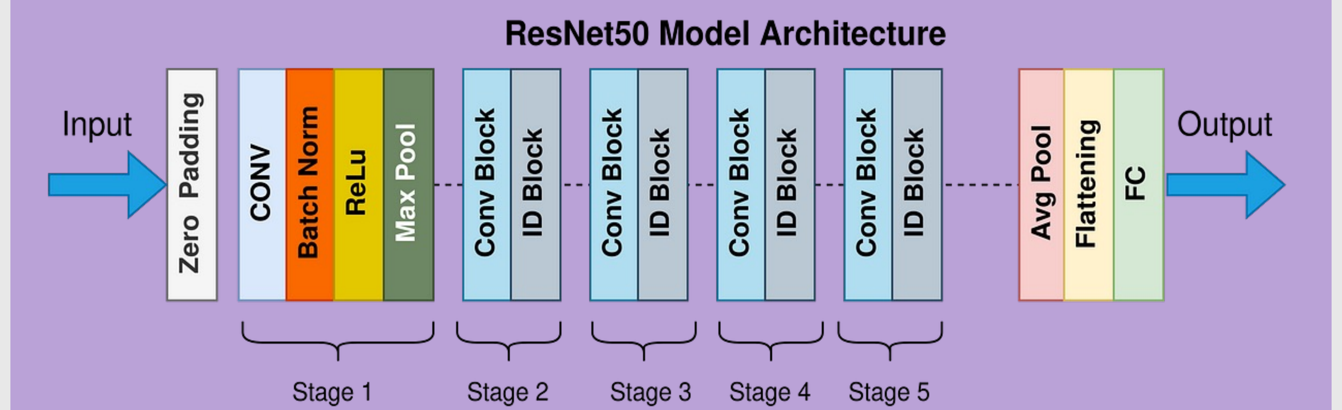

Figure 4– Standard ResNet50 architecture
https://commons.wikimedia.org/wiki/File:ResNet50.png

ResNet50 (Base): A reliable, proven image recognition architecture and the natural starting point for this pipeline. All subsequent models are measured against it. [1]

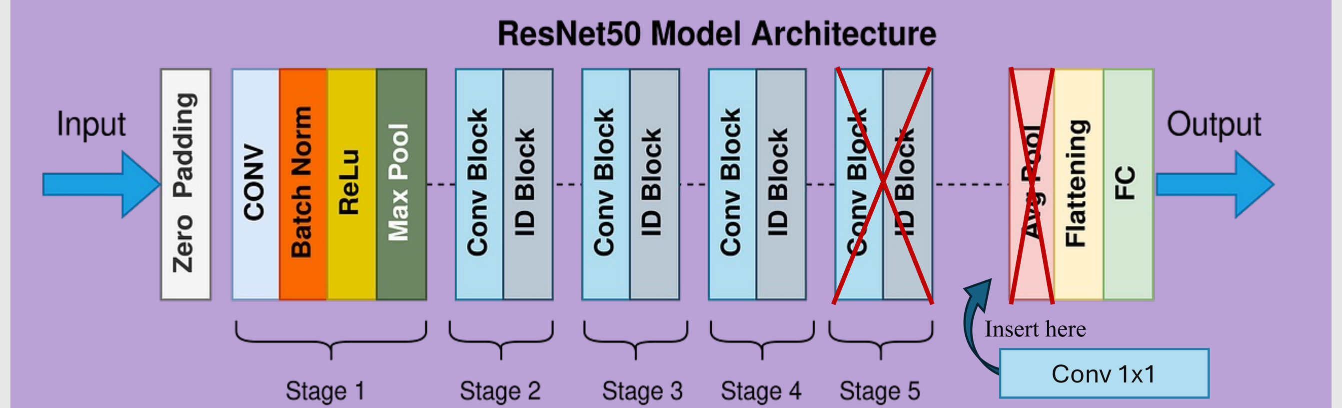

Figure 5– Truncated ResNet50 architecture

ResNet50 (Truncated): Building on the ResNet baseline, this variant was created with explainability in mind. By cutting the network after stage 4, the model retains a spatial feature grid that maps directly back to regions of the original coin image, making it possible to identify which parts of a coin are driving a match. [1]

Training Data

- Total data: 29,000+ images of Roman Republic coins
- Datasets:
  - RRC All:
    - RRC-60 [3]:
      - 12,000 images
      - 60 coin types, each with 100 obverse and reverse images (one image per coin)
    - RRCD [4]:
      - 18,000 images, reverse side (one image per coin)
      - Generally high quality, though variable
  - 9 Coins from the Sawhill Collection (final test set):
    - 9 coins, around 10 obverse and reverse images each
    - Taken manually with an iPhone under variable lighting/angles
  - RRCD and RRC-60 images are shuffled then split: 64% for training, 8% for validation and 8% for testing, the 20% of the final test set

Figure 6– Sawhill (left) vs RRCD example images

Results

To evaluate the impact of architecture on coin matching performance, we trained and tested three backbone networks (ResNet50, ConvNeXt, and Inception) each under both standard and tuned augmentation conditions using 9 coins from the JMU Sawhill Collection.

Model	RCC all Accuracy %	Sawhill Accuracy %
ResNet50 models		
• **ResNet50 Truncated (tuned aug.)**	**99.07%**	**95.48%**
• ResNet50 Truncated (standard aug.)	95.90%	93.54%
• **ResNet50 Base (tuned aug.)**	**88.87%**	**73.75%**
• ResNet50 Base (standard aug.)	99.15%	49.11%
ConvNeXt models [7]		
• ConvNeXt (standard aug.)	99.32%	64.62%
• **ConvNeXt (tuned aug.)**	**85.67%**	**85.40%**
Inception models [8]		
• Inception (standard aug.)	99.32%	24.31%
• **Inception (tuned aug.)**	**60.05%**	**35.22%**

- RCC accuracy measures how well the model matches synthetically augmented training images
- Sawhill accuracy measures real-world performance on genuinely different photos of the same coin
- Nearly all models score high on RCC, showing they learn the augmentations well
- Real-world Sawhill performance drops significantly across most models

Future Work

A key limitation of the current pipeline is the scarcity of real-world test images. The current dataset has only 9 coins from the JMU Sawhill coin collection with roughly 10 images each, captured under variable conditions. To address this, we are expanding our dataset by photographing the Sawhill Collection using a turntable, robotic arm and controlled lightbox setup, enabling consistent, high-volume image capture across many more coins.

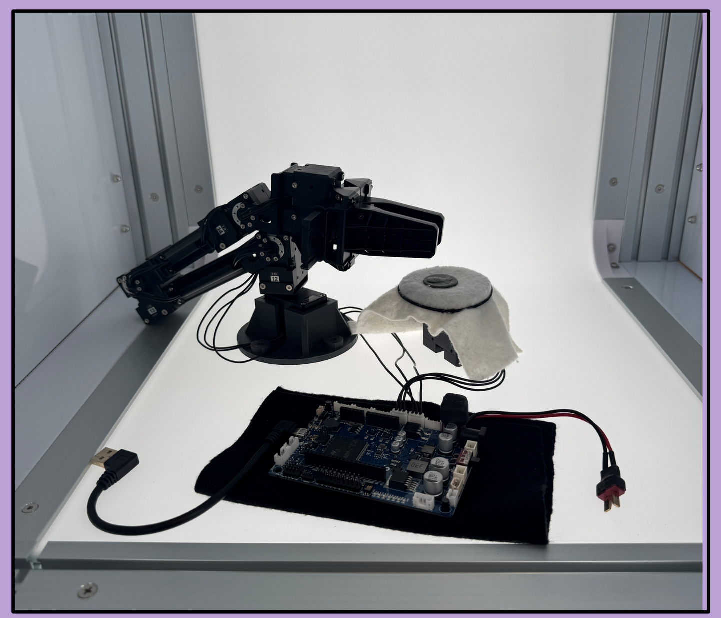

Figure 6 – Robotic arm and lightbox setup for automated coin imaging

Additionally, we are actively investigating model explainability, with the goal of identifying which regions of a coin drive a match decision, making the model's reasoning transparent and interpretable

References

1. He, K., Zhang, X., Ren, S., & Sun, J. (2016). "*Deep Residual Learning for Image Recognition.*" CVPR 2016.
2. Szegedy, C., et al. (2015). "*Going Deeper with Convolutions.*" CVPR 2015.
3. Aslan, S., Vascon, S., & Pelillo, M. (2020). Two sides of the same coin: Improved ancient coin classification using Graph Transduction Games. *Pattern Recognition Letters*, 131, 158–165.
4. Anwar, H., Anwar, S., Zambanini, S., & Porikli, F. (2021). Deep ancient Roman Republican coin classification via feature fusion and attention. *Pattern Recognition*, 114, 107871.
5. Zhuang Liu et al., "*A ConvNet for the 2020s*," arXiv:2201.03545, preprint, arXiv, March 2, 2022, https://doi.org/10.48550/arXiv.2201.03545.
6. Faria, F. A., Buris, L. H., Pereira, L. A. M., & Cappabianco, F. A. M. (2024). "*Creating Ensembles of Classifiers through UMDA for Aerial Scene Classification.*"
7. Liu, Z., Mao, H., Wu, C.-Y., Feichtenhofer, C., Darrell, T., & Xie, S. (2022). *A ConvNet for the 2020s*. CVPR 2022. https://doi.org/10.48550/arXiv.2201.03545
8. Joshi, K. D., Shah, D., Shah, V., Gandhi, N., Shah, S. J., & Shah, S. B. (2022, November). *Machine Vision Using Cellphone Camera: A comparison of deep networks for classifying three challenging denominations of Indian coins.*
9. J. Deng, J. Guo, N. Xue & S. Zafeiriou, "*ArcFace: Additive Angular Margin Loss for Deep Face Recognition*," 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 2019, pp. 4685-4694.