Deep Metric Learning for Ancient Coin Identification

JMU Department of Engineering
JMU Department of Computer Science

By: Trevor Schonbrun
Advisors: Dr. Nathan Sprague and Dr. Jason Forsyth
Department of Computer Science, James Madison University, Harrisonburg, Virginia 22807

Introduction

Establishing provenance is essential for verifying the authenticity, legality, and value of ancient coins. However, the current process of manually searching auction catalogs is slow, error-prone, and dependent on expert knowledge. To address this, we apply deep metric learning to rapidly and reliably match coins with catalog images. A key challenge is that most catalogs provide only a single image per coin, which risks overfitting to image-specific conditions such as lighting or angle. To improve robustness, we use data augmentation to simulate varied conditions and help the model focus on the true distinguishing features of each coin.

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

The Sawhill Collection

In 1976, Drs. John and Bessie Sawhill donated over 400 Ancient Greek and Roman coins to the Madison Art Collection at JMU. The collection included catalogs and notes, but no documentation of ownership history or where the coins were purchased.

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 outputs embeddings by learning a metric function where embeddings of matching instances are close (distance corresponds with 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

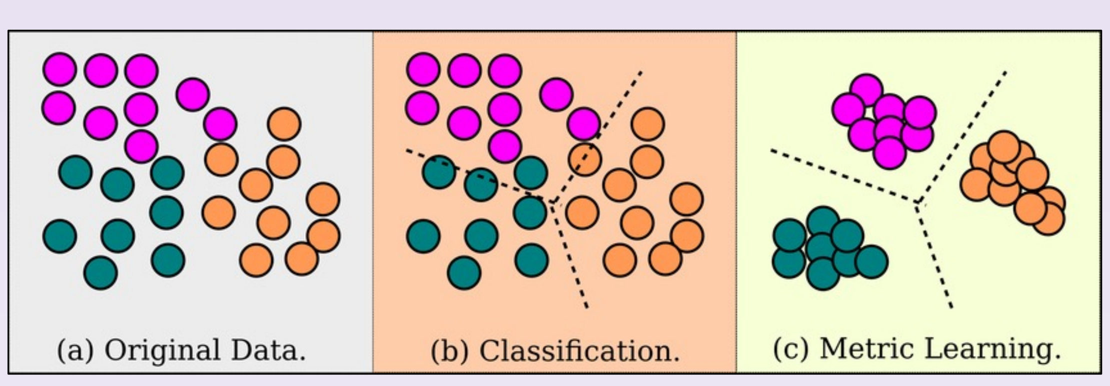

Figure 2 – Classification vs Metric learning [1]

Training Data

- Total data: 29,000+ images of Roman Republic coins
- Datasets:
  - RRC-60 [2]:
    - 12,000 images
    - 60 coin types, each with 100 obverse and reverse images (one image per coin)
  - RRCD [3]:
    - 18,000 images, reverse side (one image per coin)
    - Generally high quality, though variable
  - 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: 80% for training, 10% for validation and testing

Model Layout

- ResNet50 backbone [5]:
  - Pretrained deep convolutional neural network
  - Final softmax classification layer replaced with a 512-d layer with ArcFace loss to create embeddings

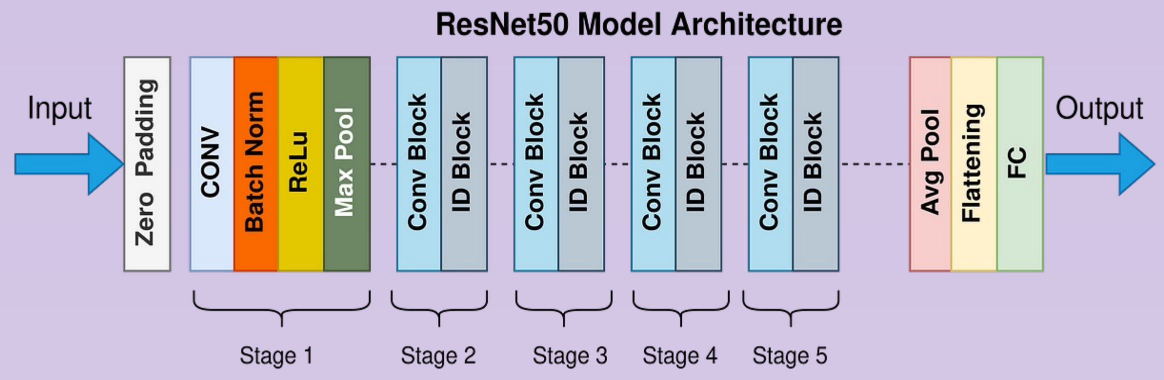

Figure 3 – Standard ResNet50 architecture
https://commons.wikimedia.org/wiki/File:ResNet50.png

- Softmax cross-entropy loss:
  - Softmax converts logits (raw outputs) into a probability distribution
  - The loss function is:

$$L = -\frac{1}{N}\sum_{i=1}^{N}\log\frac{e^{z_{y_i}}}{\sum_{j=1}^{n} e^{z_j}}$$

where N is the batch size, n is the class number, z_{yi} is the logit of the true class, z_j is the logit for other classes

- Additive Angular Margin (ArcFace) loss [4]: similar to softmax, but…
  - Normalizes embedding features and weights, projecting them onto a hypersphere of radius s.
  - Predictions depend only on the angle between an embedding and weight.
  - Adds an angular margin.
  - Increases intra-class compactness, inter-class separation.

- The loss function is:

$$L = -\frac{1}{N}\sum_{i=1}^{N}\log\frac{e^{s\left(\cos\left(\theta_{y_i}+m\right)\right)}}{e^{s\left(\cos\left(\theta_{y_i}+m\right)\right)}+\sum_{j=1,j\neq y_i}^{n} e^{s\left(\cos\left(\theta_j\right)\right)}}$$

where θ_{yi} is the logit of the true class, θ_j is the logit for other classes, s is the hypersphere radius, m is additive angular margin penalty

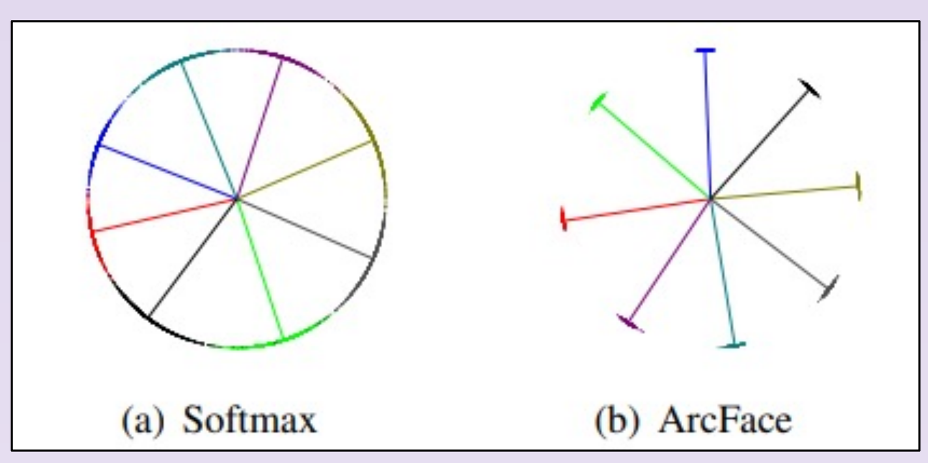

Figure 4 – Visualization of softmax and ArcFace. Dots represent samples, while lines are class centers. [4]

Data Augmentation

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

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

Figure 5 – Example augmentations

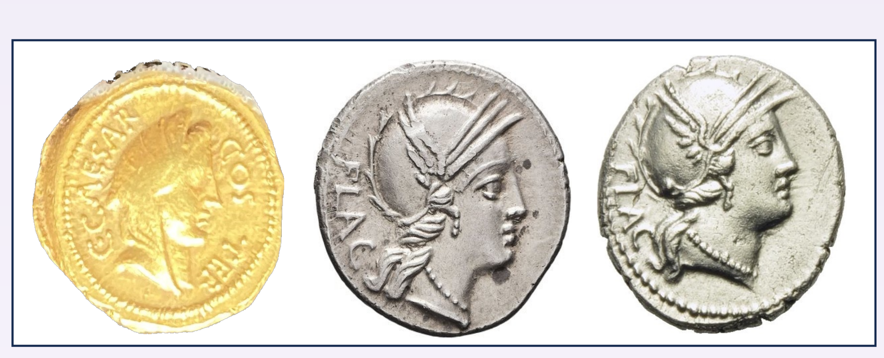

Figure 6 – Sawhill (left) vs RRCD example images

Results

After training, our model reached 99.15% accuracy on the holdout test set from RRC-60 and RRCD. Out of 2954 images, 25 were incorrectly identified. The accuracy of the model greatly increased after training.

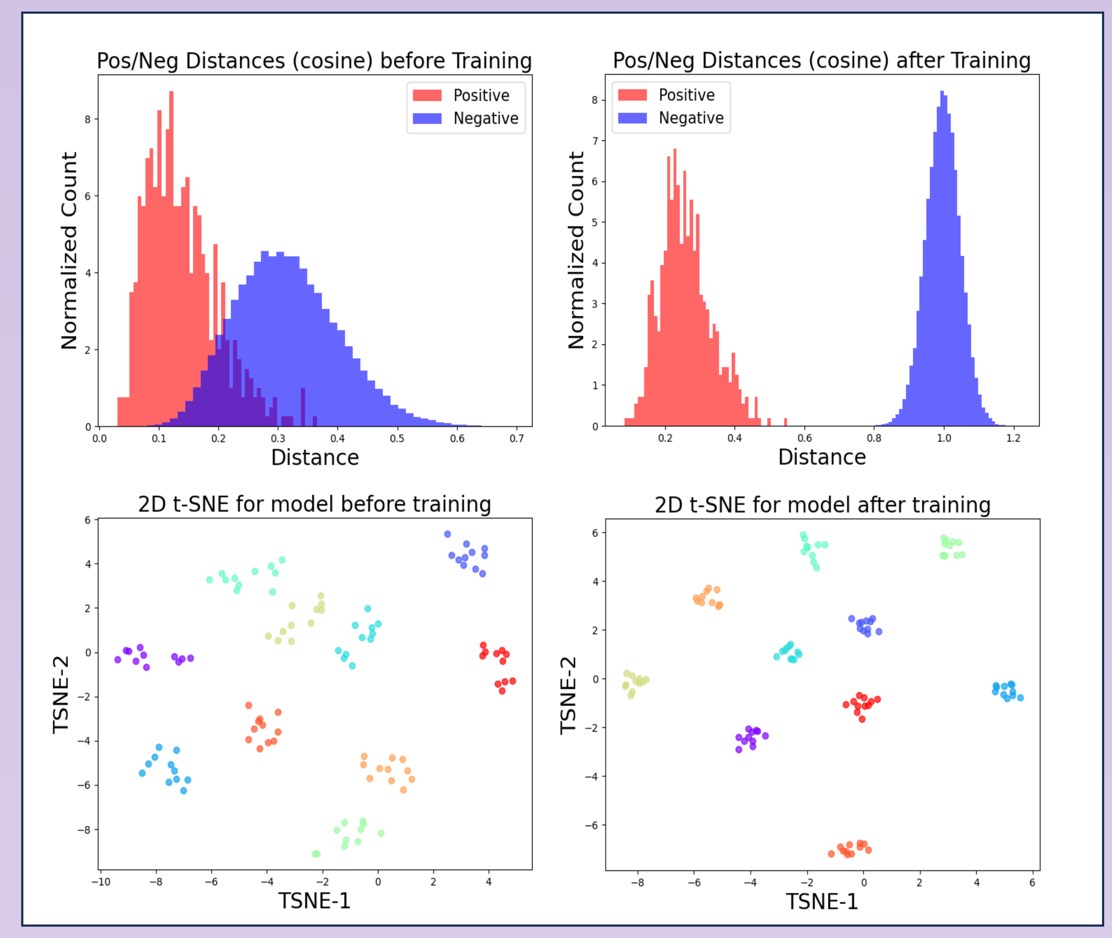

Figure 7 – Distance histogram (top) and t-SNE (bottom) on RRC-60 and RRCD images before and after training

Figure 8 – Visualization of nearest neighbors after training

Sawhill Results

To run a final test on our data, we put the Sawhill images through the model before and after training. On this data, the increase in accuracy is visible, but it is not as profound.

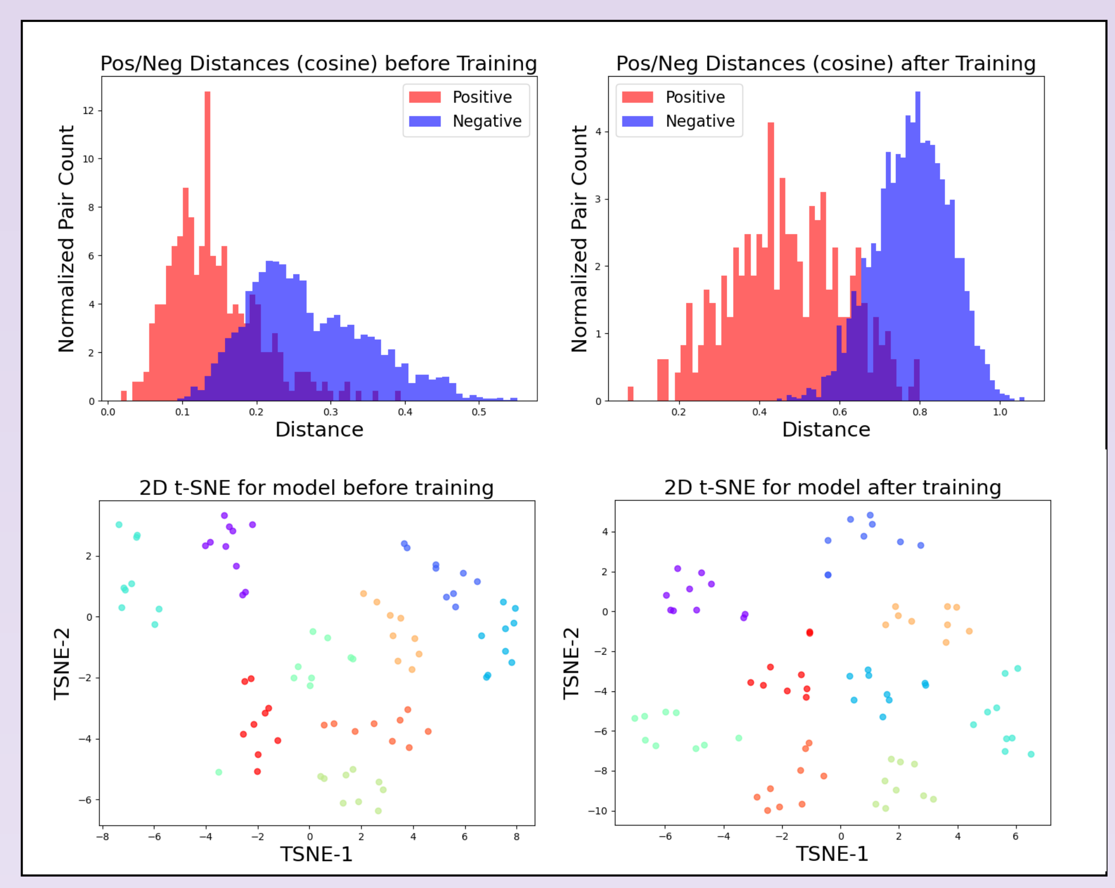

Figure 9 – t-SNE on Sawhill images before and after training

Future Work

- Data preprocessing:
  - RRC-60 and RRCD may have unwanted duplicate images
- Explore different augmentations to improve the robustness and generalizability of the model
- Visualization tool showing key features used by the model in coin identification

REFERENCES

[1] 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."

[2] 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.

[3] Anwar, H., Anwar, S., Zambanini, S., & Porikli, F. (2021). Deep ancient Roman Republican coin classification via feature fusion and attention. *Pattern Recognition*, *114*, 107871.

[4] 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.

[5] He, K., Zhang, X., Ren, S., & Sun, J. (n.d.). *Deep Residual Learning for Image Recognition*. Retrieved September 27, 2025, from http://image-net.org/challenges/LSVRC/2015/