Sneaker Authenticity Detection :

A Multimodal Vision Transformer Approach

CS 441: Final Project Presentation

Team Members *:*

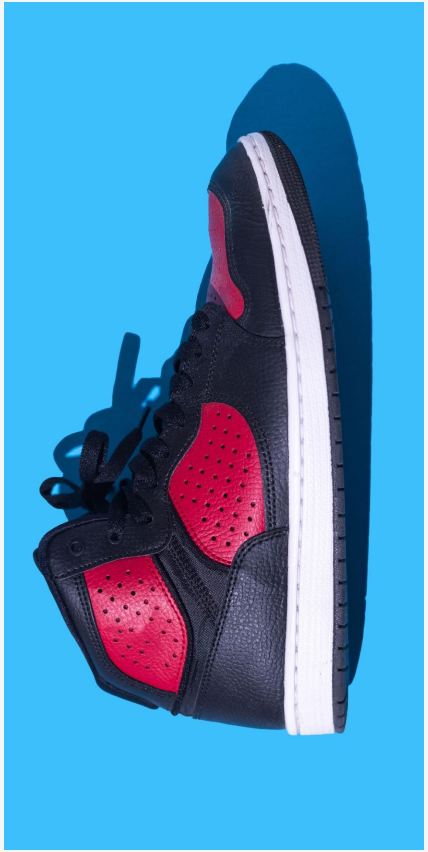
- Praveen Rajendar Sundar (pr36)
- Bhoomika Narasimhulu (bhoomik2)
- Akash Elumalai (akashe2)
- Abhirup Chakraborty (abhirup4)

Motivation & Problem Statement :

The Challenge of Counterfeits :

Counterfeit sneakers are visually similar to authentic ones.
Manual verification is slow and error-prone.

- **Goal:** Distinguish between authentic and fake sneakers.
- **Complexity:** High-quality counterfeits look visually similar to authentic products.
- **Solution:** We cannot rely on images alone. We need a holistic view:
  1. **Visual Details:** Stitching, logo shape, materials.
  2. **Market Context:** Price points (too cheap to be true?) and brand-specific patterns.



Dataset Overview :

Images sourced from multiple sneaker brands.

Labels :

- 0 => Fake
- 1 => Authentic

Dataset collection :

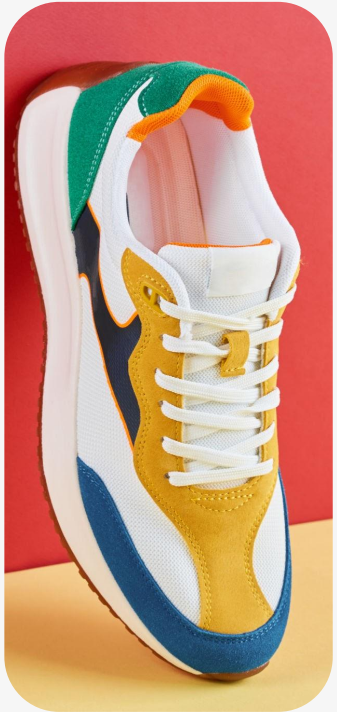
1. Manually via Google images, Reddit, Counterfeit product websites such as ioffer.com and so on.
2. Most of the authentic images are scraped via a python script

A	B	C	D	E	F
sln	Image Name	Brand	Price	Authentic	
1	1.png	Jordan	29	0	
2	2.png	Jordan	35	0	
3	3.png	Jordan	49	0	
4	4.png	Jordan	50	0	
5	5.png	Jordan	37	0	
6	6.png	Jordan	29	0	
7	7.png	Jordan	29	0	
8	8.png	Jordan	49	0	
9	9.png	Jordan	49	0	
10	10.png	Jordan	59	0	
11	11.png	Jordan	3000	0	
12	12.png	Jordan	49	0	
13	13.png	Jordan	69	0	
14	14.png	Jordan	98	0	
15	15.png	Jordan	59	0	
16	16.png	Jordan	79	0	
17	17.png	Jordan	49	0	
18	18.png	Jordan	69	0	
19	19.png	Jordan	70	0	
20	20.png	Jordan	50	0	
21	21.png	Jordan	36	0	
22	22.png	Jordan	29	0	
23	23.png	Jordan	49	0	
24	1.jpg	Jordan	130	1	

+ ≡ Nike Adidas Reebok Test Jordan Puma

Data Pipeline :

1. **Data Source:** Scraped images and metadata organized by brand (Adidas, Jordan, Nike, Puma, Reebok).
2. **Structure:**
    a. Authentic vs Fake folders.
    b. Metadata: Brand Name, Price, Authenticity Label.
3. **Preprocessing:**
    a. **Images:** Resized to 224x224, Normalized (ImageNet stats), Augmented (Flip, Rotation, Color Jitter).
    b. **Price:** Log-transformed (log1p) and Standardized to handle wide price ranges.
    c. **Brand:** Encoded as categorical integers.



Model Architecture :

Multimodal Vision Transformer (ViT)

We utilize a fusion architecture with three input paths:

1. **Image Path (ViT-Base) :**
    a. Pretrained on ImageNet.
    b. Partially frozen (first 2 layers) to retain low-level features while adapting high-level features.
    c. Outputs a 512-dim embedding.
2. **Brand Path:**
    a. Learnable Embedding (64-dim) captures brand-specific characteristics.
3. **Price Path:**
    a. MLP projection (64-dim) contextualizes price relative to the market.

Fusion: Features are concatenated and passed through a dense Authenticity Head.

Training Methodology :

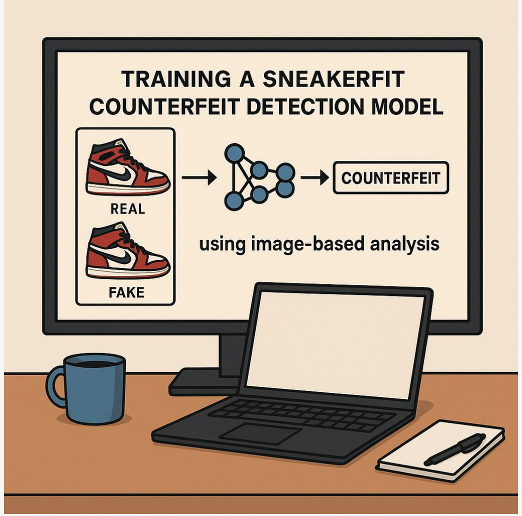
Multi-Task Learning

To improve robustness, we train on three objectives simultaneously:

1. **Authenticity Loss (Main):** Binary Cross Entropy (BCE) to predict Real vs. Fake.
2. **Auxiliary Brand Loss:** Cross Entropy loss ensuring the model preserves brand identity in the image features.
3. **Contrastive Loss:** Supervised contrastive loss to cluster similar brand representations together.

Optimization:

- **Optimizer:** AdamW (Weight Decay 1e-4).
- **Scheduler:** Cosine Annealing with Warmup.
- **Regularization:** Dropout, Early Stopping.



Results :

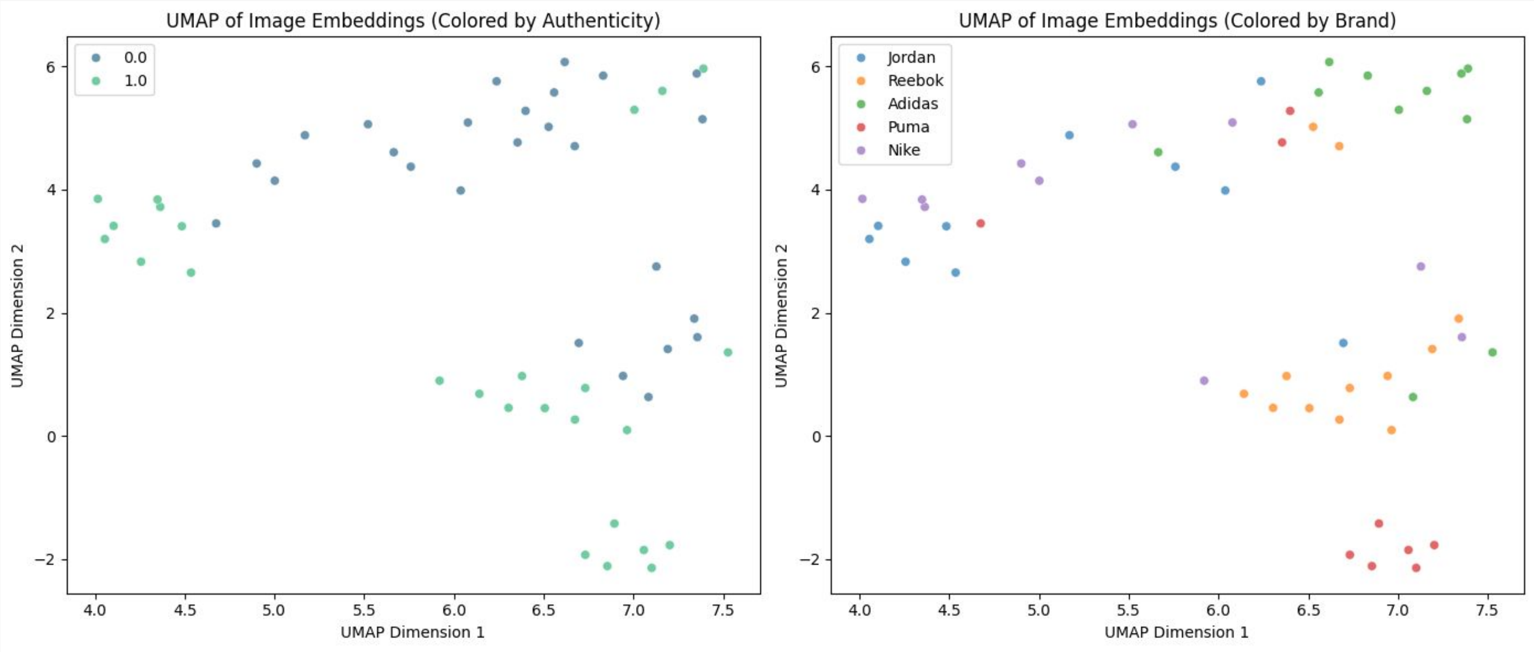
Performance Metrics :

- **Validation Accuracy:** ~96% (Best Epoch)
- **Test Set Accuracy:** 86.36%
- **Test AUC:** 0.8824

Key Findings:

- The model successfully balances visual cues with price context.
- Partial freezing of the ViT backbone allowed effective transfer learning on a small dataset (~250 images).
- High AUC indicates the model is robust in ranking authentic items higher than fakes, even if the decision threshold needs tuning.

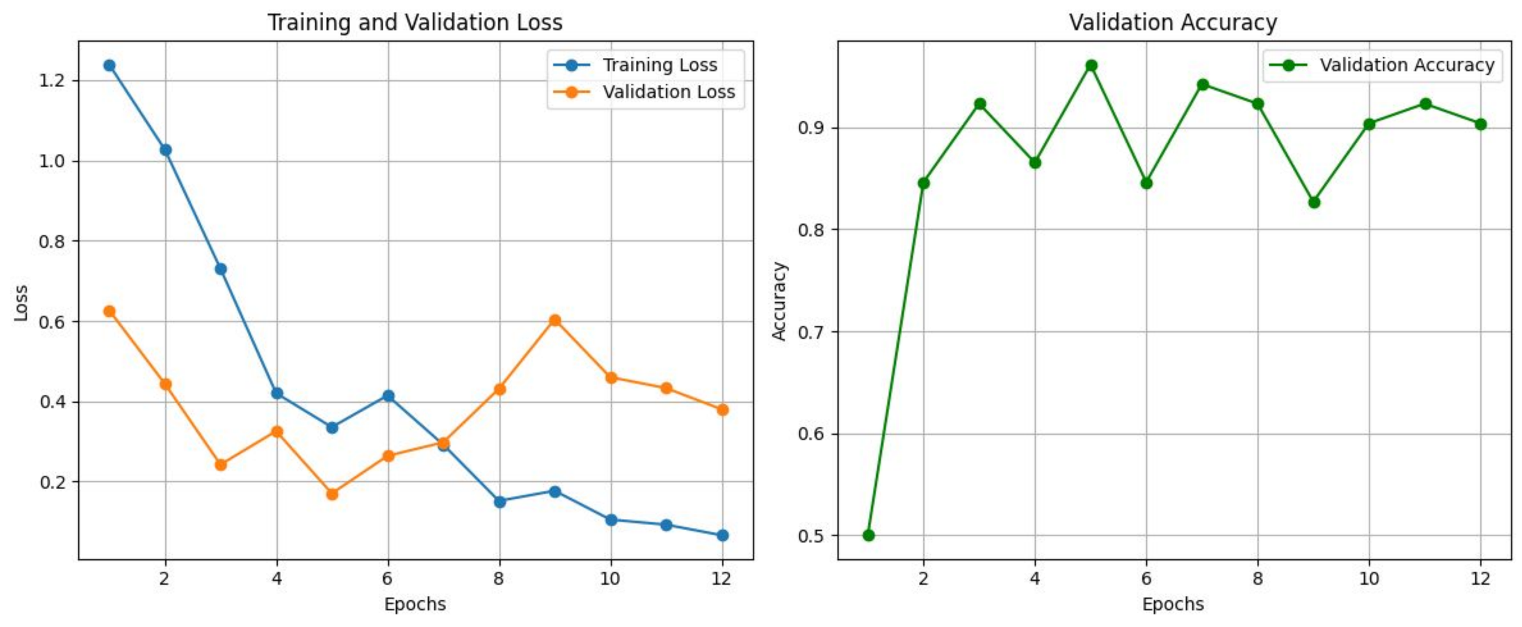
Feature Visualization (UMAP) :



Feature Visualization (UMAP) :

- **Method:** UMAP projection of the 512-dimensional image embeddings into 2D space.
- **Observation (Brand):** The model forms distinct clusters for different brands (right plot), confirming it has learned brand-specific visual characteristics.
- **Observation (Authenticity):** There is a visible separation between 'Authentic' and 'Fake' samples (left plot), though the complex decision boundary reflects the difficulty of the task.

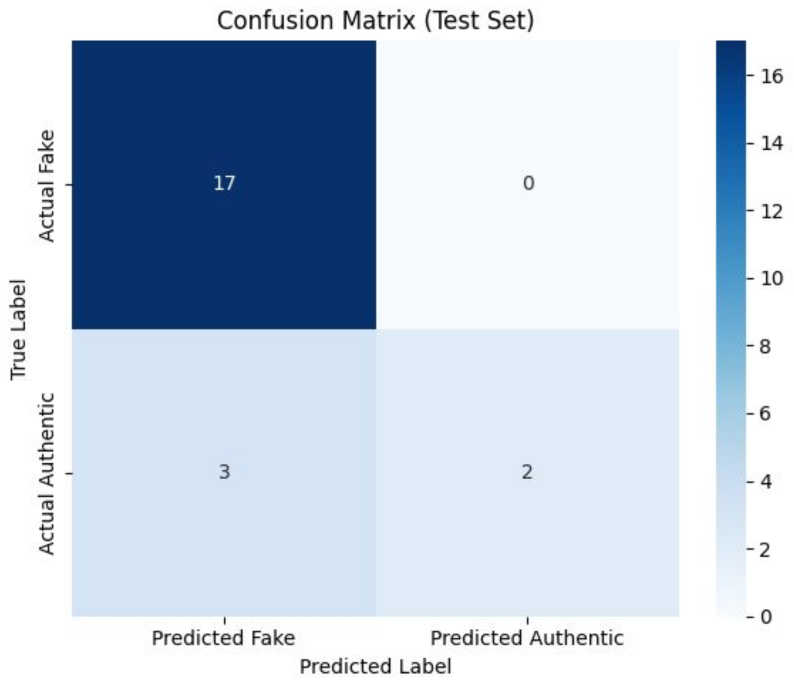
Training Dynamics :

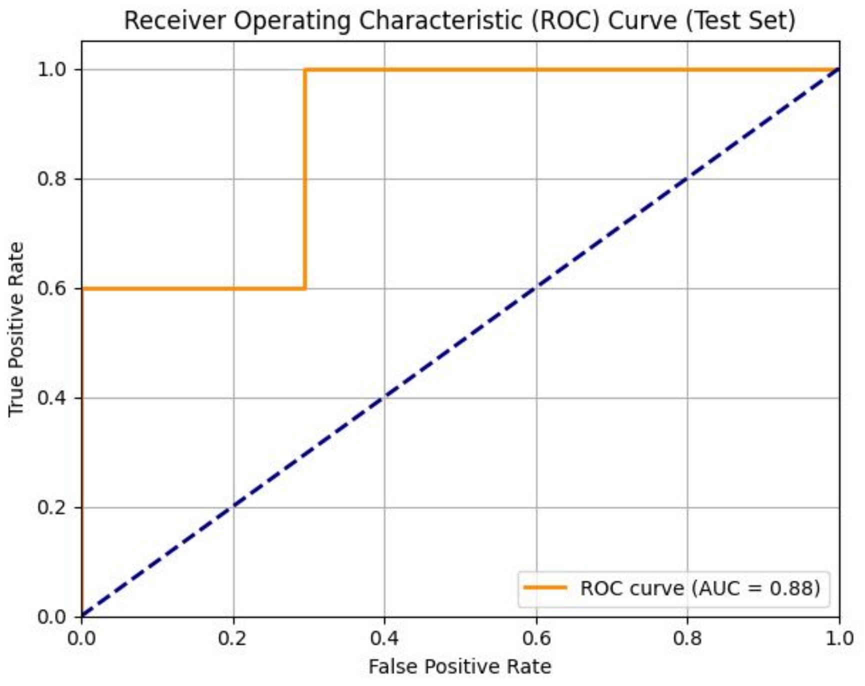


Training Dynamics :

- **Accuracy:** Validation accuracy peaked at ~96% during training.
- **Loss Analysis:** Training loss consistently decreased, indicating effective learning. Validation loss improved initially but then stabilized, triggering Early Stopping at Epoch 12 to prevent overfitting.
- **Takeaway:** The model converges quickly (within 5 epochs) to a high-performance state.

Model Evaluation (Test Set) :





Model Evaluation (Test Set) :

- **Metric:** Test Accuracy 86.36% with an AUC of 0.88.
- **Confusion Matrix Insight:**
  - **0 False Authentics:** The model correctly identified all fakes (17/17). It did not let any counterfeit pass as real.
  - **3 False Fakes:** It is slightly conservative, flagging 3 authentic pairs as suspicious. In a fraud detection context, this is the preferred behavior (better to double-check a real item than miss a fake).
- **Conclusion:** The system is highly reliable for screening, acting as a strict gatekeeper against counterfeits.

Thank you !