

# Image Embedding for Product Labeling

Couture Vision

January 15, 2025

# Project Overview

- ▶ We utilized Microsoft's ResNet50 and Google's ViT to extract embeddings.
- ▶ Cosine similarity is used to find the best matches between images.
- ▶ We preprocess images using `rembg` to remove backgrounds, as test images (shop pictures) differ from the clean DAM images.
- ▶ Embedding extraction assists in labeling products by suggesting potential matches.

# Embedding Extraction Pipeline

- ▶ Use ResNet50 and ViT to extract image embeddings.
- ▶ Employ cosine similarity to rank matches.
- ▶ Preprocess test images with rembg to isolate objects from noisy backgrounds.

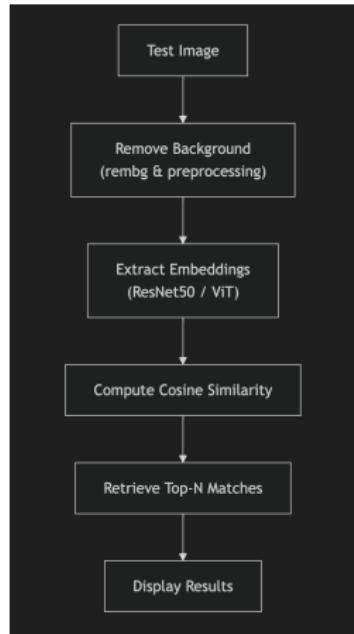


Figure: Pipeline overview

# Embedding Extraction Pipeline

Original Image



Extracted Object



Figure: Background removal using rembg, scaling and center-cropping

# Good Classification Example

- ▶ Original image and extracted object.
- ▶ Top 5 matches with high similarity scores.



Match 1  
Sim: 0.7321



Match 2  
Sim: 0.7247



Match 3  
Sim: 0.6787



Match 4  
Sim: 0.6785



Match 5  
Sim: 0.6712



## Bad Classification Example

- ▶ Challenges: reflections on the glass adversely affecting background removal and embeddings.



## Enhancing the Approach

- ▶ Use **TRELLIS** for generating 3D models and performing data augmentation with various object poses.
- ▶ Explore **Siamese Networks** to refine embeddings by focusing on important features.



Figure: Input for TRELLIS

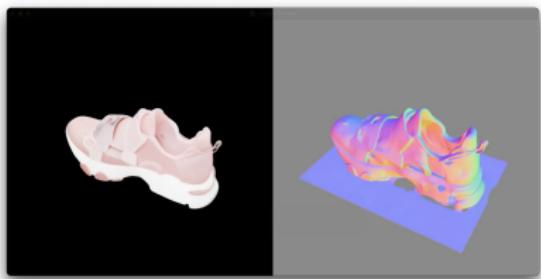


Figure: TRELLIS generated 3D model

## Web Interface with Gradio

- ▶ Develop a quick UI using Gradio.
- ▶ Users can upload a picture and get top-X matches.
- ▶ This tool can aid in manual labeling and refine embeddings.

## ViTMSN Model

- ▶ Considering the ViTMSN approach for improved performance.
- ▶ Joint-embedding architecture from *Masked Siamese Networks for Label-Efficient Learning*.
- ▶ Promising for low-shot and extreme low-shot regimes.