**Thuli Studios - End-to-End Fashion Accessory Detection Pipeline**

**1. Introduction**

The objective was to develop a computer vision system capable of detecting and tracking:

* Rings on human hands
* Earrings on ears
* Dresses on the body

This system should work on real-world videos with diverse hand movements and different backgrounds.

**2. Dataset Creation**

**2.1 Sources**

We curated and combined three different sources:

* **Ring Dataset**: Roboflow and custom Google-scraped images showing rings on hands.
* **Earring Dataset**: Roboflow public datasets.
* **Dress Dataset**: Fashionpedia and Roboflow sources.

**2.2 Merging**

* Standardized all labels into YOLOv8 format.
* Remapped class indices:
  + 0 → Ring
  + 1 → Earring
  + 2 → Dress
* Merged images and labels into a unified folder structure:

multi-fashion-dataset/

├── train/images

├── train/labels

├── valid/images

├── valid/labels

└── data.yaml

**3. Model Training**

**3.1 Model Selection**

* **YOLOv8n** (nano model) was chosen for its balance between speed and accuracy.

**3.2 Training Configuration**

* **Epochs**: 50
* **Batch Size**: 16
* **Image Size**: 640x640
* **Optimizer**: Auto-selected by Ultralytics
* **Augmentations**: Mosaic, random flipping, random brightness, random scale

**3.3 Training Environment**

* Python 3.10
* Ultralytics 8.3.111
* CUDA: NVIDIA RTX 3070 Ti GPU

**4. Inference Pipeline**

**4.1 Main Script**

* main.py reads input video.
* Applies trained YOLOv8 model frame-by-frame.
* Annotates detections with class labels and confidence scores.
* Saves output video with annotations.

**4.2 Classes Detected**

* **Ring**
* **Earring**
* **Dress**

**4.3 Post-Processing**

* Filtering based on confidence threshold (>25%).
* Bounding box annotations with frame counters.

**5. Observations and Challenges**

**5.1 What Worked Well**

* Earrings and dresses were reliably detected under diverse conditions.
* Dataset merging provided a balanced class distribution.

**5.2 Challenges**

* Rings required extremely high-quality, close-up hand images.
* Small object detection (rings) is inherently harder at lower resolutions.
* Background clutter sometimes affected small object recall.

**6. Future Work**

**6.1 Data Improvements**

* Collect more ring-on-hand videos under diverse lighting and skin tones.
* Introduce hands in motion blur scenarios during training.

**6.2 Model Enhancements**

* Test larger YOLOv8 variants (YOLOv8m or YOLOv8l) for higher precision.
* Explore using DETR-based models or hybrid pipelines for rings.

**6.3 3D Modeling (Optional Extension)**

* Develop a Blender-based 3D rig of Anna's hand.
* Attach virtual rings and earrings realistically.
* Enable try-on AR visualizations in future releases.

**7. Conclusion**

I successfully designed, built, and tested an end-to-end fashion accessory detection system. The solution is modular, scalable, and can be extended into a full production-grade virtual try-on experience with minor future enhancements.