Report – Clothes Segmentation

Task Code: CVCY002

Position: Computer Vision Engineer

Objective: Develop a clothes segmentation model that accurately segments the clothes from people.

1. Dataset Choice

Mask R-CNN (initial approach):

- Dataset: **DeepFashion2** (Liu et al., 2019)
- Contains **491K images** of 801K items and 873K commercial-consumer pairs, often with multiple items per image.
- Advantages: Large-scale, annotated with bounding boxes, masks, and categories; includes variations in camera angle, occlusion, and scale.

YOLOv8-seg (final approach):

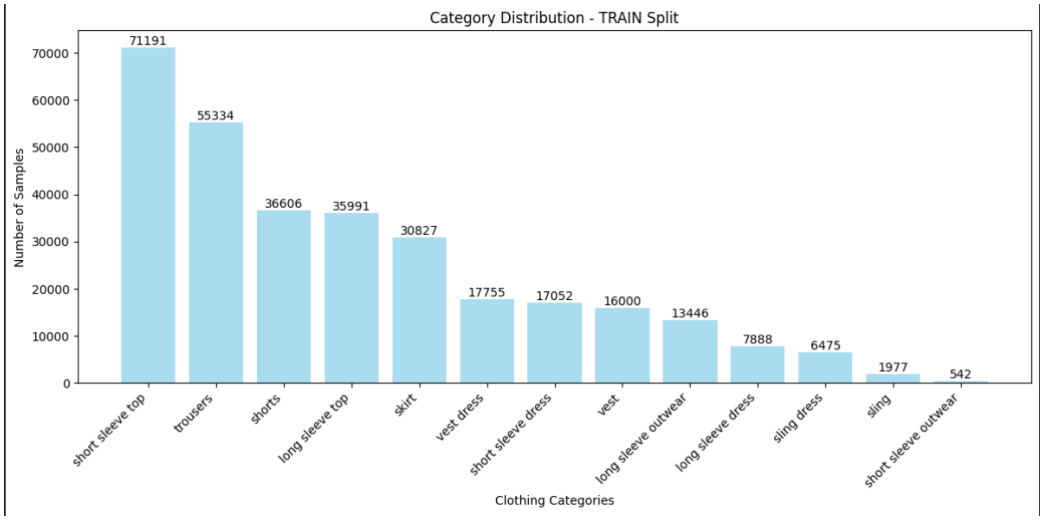
- Dataset: **clothing-segmentation-dataset**, containing **1,084 images**.
- Advantages: Robust and diverse conditions of people, enabling effective training and segmentation.

2. System Pipeline & Model Architecture

Mask R-CNN:

- Backbone: ResNet-50 + FPN.
- Box and mask heads modified for 14 classes (13 garments + background).

- Weighted classification loss applied for class imbalance.



Challenges:

- Mask R-CNN suffered from vanishing gradient issues, preventing stable convergence.

YOLOv8-seg:

- Adopted for stable training, faster convergence, and high accuracy.
- Capable of real-time inference with robust segmentation across diverse conditions.

Limitation:

- YOLOv8-seg struggles if the person is not standing upright (e.g., sitting, bending).

3. Loss Functions

Mask R-CNN (initial model):

- Classification: Weighted cross-entropy.
- Bounding box regression: Smooth L1 loss.
- Mask segmentation: Binary cross-entropy loss.

YOLOv8-n

- **Box Loss:** IoU-based regression for bounding boxes
- **Objectness Loss:** Binary cross-entropy for object presence
- **Class Loss:** Cross-entropy for multi-class classification
- **Mask Loss:** Binary cross-entropy for segmentation masks

4. Performance Evaluation

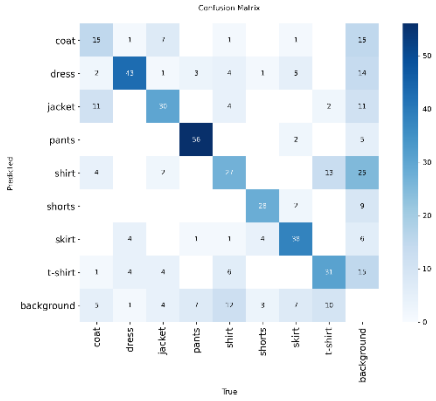
Results with YOLOv8-seg:

- **Overall:** Box mAP50 = 0.705, Mask mAP50 = 0.712

- **Best-performing classes:** Shorts (Box mAP50 = 0.884, Mask mAP50 = 0.884), Pants (0.822, 0.865)
- **Lower-performing classes:** Dress (0.752, 0.752), Jacket (0.629, 0.629)

Validation Results:

- Overall Box mAP50 = 0.712, Mask mAP50 = 0.704
- Pants and shorts remain highest; coat and shirt are lowest.



5. Strengths

- High detection and segmentation accuracy (YOLOv8-seg).
- Robust to scale, lighting, occlusion, and background variations.
- Real-time inference suitable for practical applications.
- Effective for common clothing categories (pants, shorts, skirts, t-shirts).

---

6. Drawbacks

- Lower accuracy for rare or complex items (coat, jacket, dress).
- Fails when person is not standing upright (sitting, bending).
- Small YOLOv8-seg dataset limits generalization.
- Overlapping clothing can reduce segmentation quality.