**Detecting and Classifying People Using Detectron2**

**1. Introduction**

Object detection and classification are fundamental tasks in computer vision, with applications ranging from surveillance to automated content filtering. This project focuses on training a **Mask R-CNN** model using Detectron2 on the **LV-MHP-v2 dataset** to detect and classify people into three categories: **man, woman, and child**, while ensuring an additional modification for privacy protection. Specifically, for instances labeled as "woman," the model must be trained to mask the entire body while excluding the **face and hands** from the segmentation mask.

To achieve this, we fine-tuned a **Mask R-CNN model** on a carefully curated dataset, implementing custom preprocessing steps to modify segmentation masks before training. This report outlines the challenges encountered in setting up the dataset, the data preparation pipeline, training details, and the evaluation results.

**2. Dataset Preparation**

**2.1 Challenges in Setting Up the Dataset and Detectron2**

Setting up the dataset and configuring Detectron2 for training was not straightforward. The **LV-MHP-v2 dataset** contains multiple annotations per image, requiring extensive preprocessing to ensure compatibility with Detectron2’s format. Additionally, the dataset lacked predefined **gender labels**, necessitating a manual effort to categorize 700 images with gender-based annotations. Detectron2 itself required specific dependencies and configurations, which needed to be resolved before successful integration.

**2.2 Labeling and Dataset Splitting**

To ensure an effective evaluation, we labeled and categorized **700 images**, dividing them into:

* **500 images for training**
* **200 images for validation**

These images contain annotated segmentation masks, bounding boxes, and gender classifications. The primary objective is to train the model to correctly detect and classify people while ensuring proper handling of segmentation masks for women.

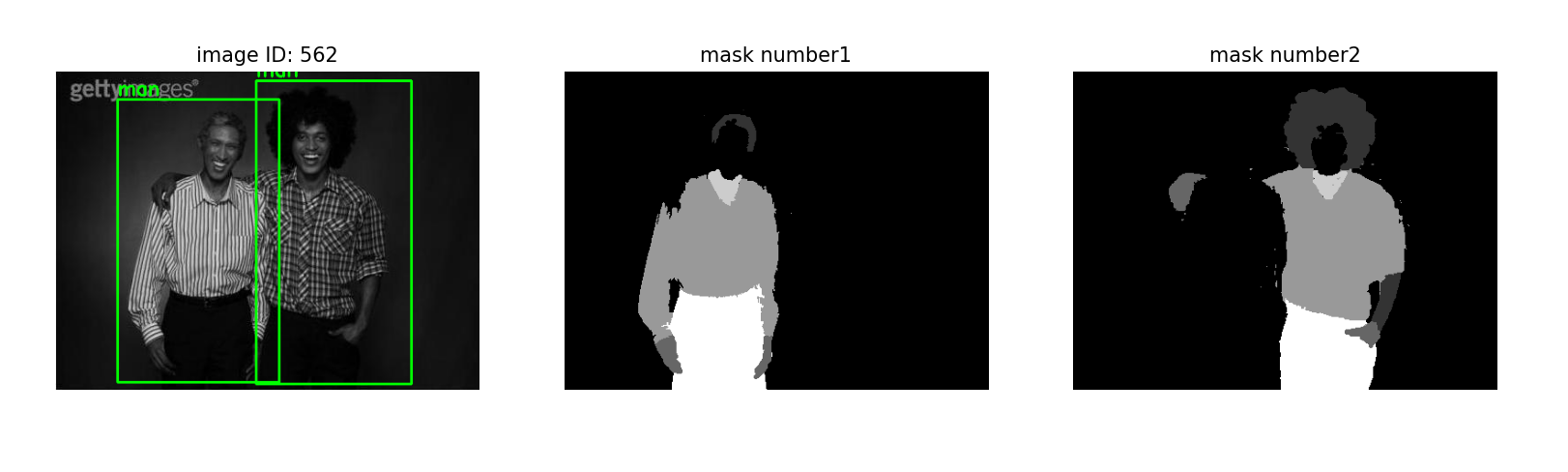
**2.3 Modifying Segmentation Masks**

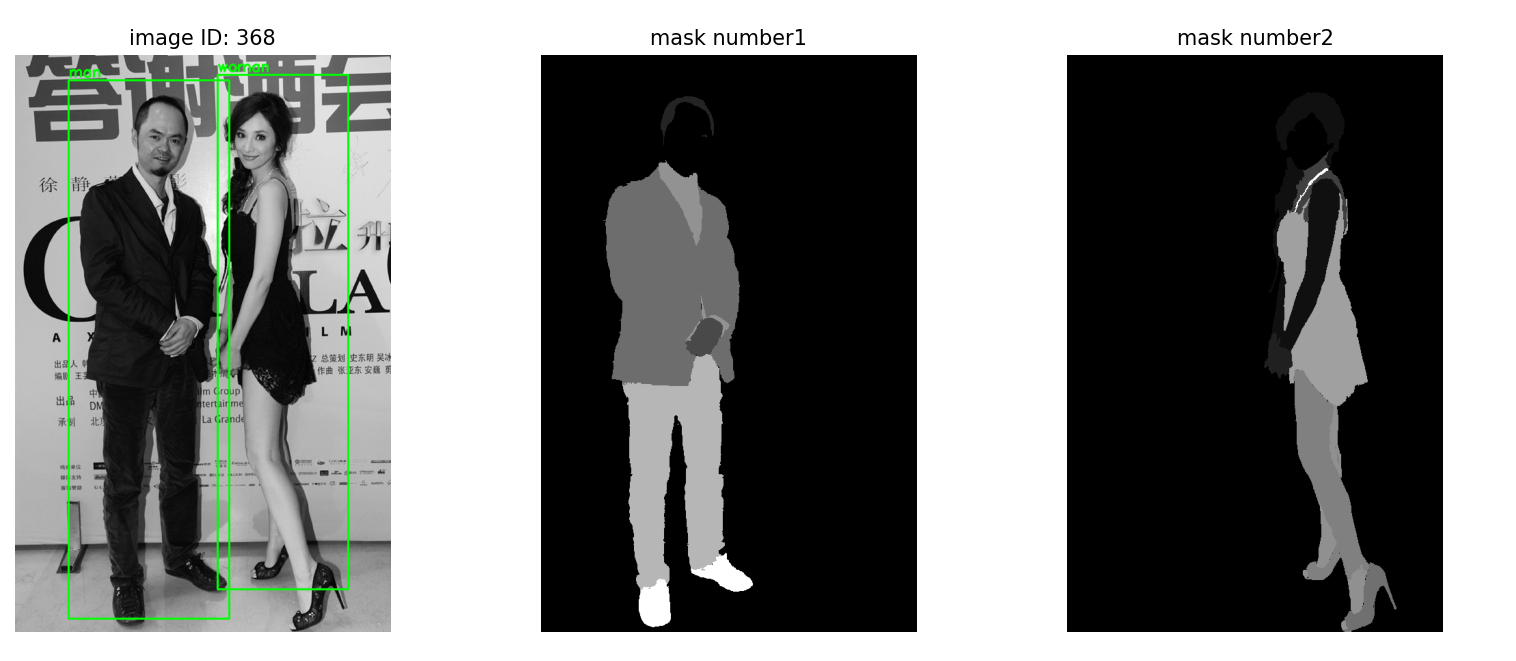
A key aspect of this project was modifying segmentation masks for instances labeled as **"woman."** The dataset originally included full-body masks, but to align with the project’s objectives, we ensured that:

* The **face and hands** (previously labeled under segmentation keys 3, 7, and 8) were **excluded** from the mask.
* The rest of the body remained masked to preserve privacy.

This preprocessing step ensures that the model learns to classify and segment women while maintaining an exclusion for the face and hands.

**2.4 Sample Images Before Training**

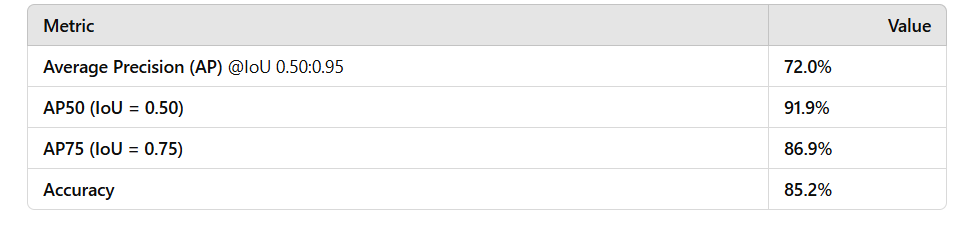
Below, we present a selection of images from the dataset. Some images include segmentation masks, while others showcase bounding boxes with gender classifications.



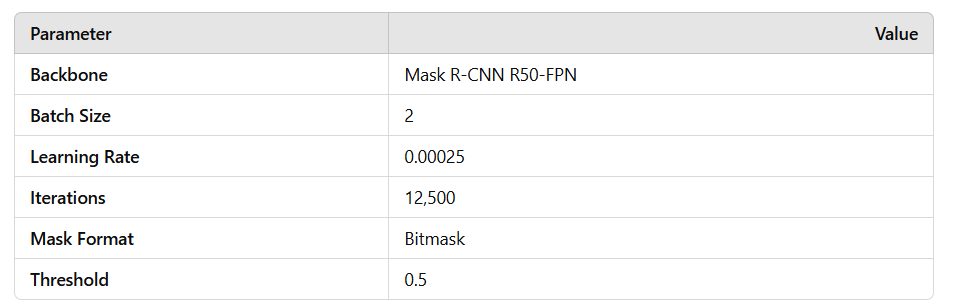
## ****Model Evaluation and Comparison****

After training two different models for **gender classification with and without masking**, we conducted a detailed evaluation to compare their effectiveness. The evaluation considered **both numerical performance metrics** and **visual results** for the first model.

### ****Model 1: Classification Without Masking****

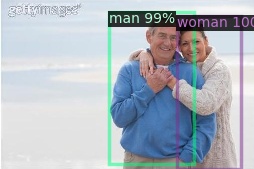
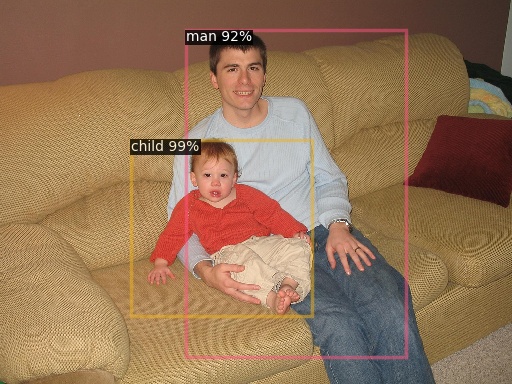
****The **first model** was trained **without any modifications to the segmentation masks**, meaning that it could use **full-body features** to distinguish between categories (**man, woman, and child**).

* **High classification accuracy (85.2%)** confirms that the model **effectively distinguishes between categories**.
* The model shows **strong precision across all IoU thresholds**, indicating **good bounding box alignment and classification**.

******

#### **Sample Visual Results**

Below are sample images showing:  
 **Bounding boxes with gender classification before training**  
 **Final detections from the trained model on the validation dataset**

****



These visual samples confirm that the model **successfully classifies individuals** in most cases. However, some **misclassifications still exist**, especially in cases where **features overlap or dataset limitations occur**.

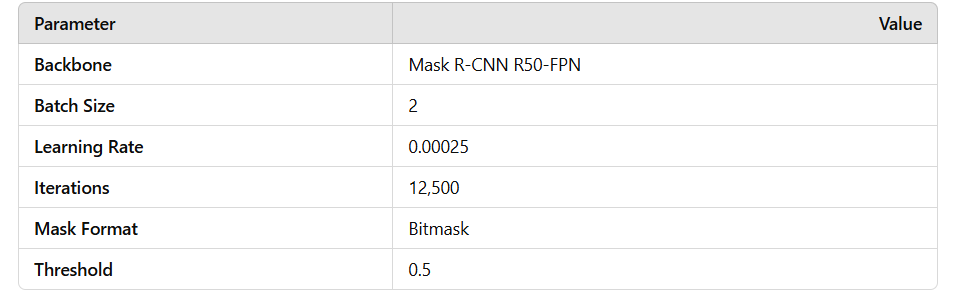
### ****Model 2: Classification With Masking****

The **second model** was trained with **a masking strategy that occludes body regions** for the "woman" category, leaving only **hands and faces visible**. The goal was to **test the impact of removing full-body features** on classification accuracy.



#### **Drastic accuracy drop (43.7%)** suggests that **gender classification heavily relies on full-body features**.

* **Bounding box precision is significantly lower** than in the first model.
* **Segmentation results were unreliable**, with **incorrect detections and gender classifications**.



#### **Why Did Performance Drop?**

1. **Feature Removal** 🏷️
   * **The full-body silhouette is essential for gender recognition.**
   * **By masking the body (except hands and face), key distinguishing features were lost.**
2. **Insufficient Training Data**
   * The dataset contained **only 700 labeled images for training**.
   * This likely **limited the model’s ability to generalize classification under occlusion.**
3. **Segmentation Quality Issues**
   * The masks may have contained **inaccuracies in distinguishing between hands, face, and occluded regions**.

### ****Final Observations****

* **Model 1 (without masking) achieved strong accuracy (85.2%)** and **high average precision across multiple IoU thresholds**.
* **Model 2 (with masking) struggled significantly (43.7% accuracy), confirming that gender classification depends on full-body features.**
* **Bounding box detection was reliable in Model 1 but not in Model 2.**
* **No visual samples for Model 2 will be provided** since **the detections were inaccurate**, and **there is no time left for additional improvements**.