University of Tripoli – Faculty of Electric and Electronics Engineering Department

---

أحمد عبد القادر بركة
2200207979

Abstract

This report explores a computer vision problem involving instance segmentation and gender classification. The LV-MHP-v2 dataset is used, which contains images of people along with parsing annotations for each person in the image. These parsing annotations mask the entire body except for the face. Additionally, a JSON file provided by the professor includes the paths to each image, along with the corresponding bounding boxes and parsing annotations.

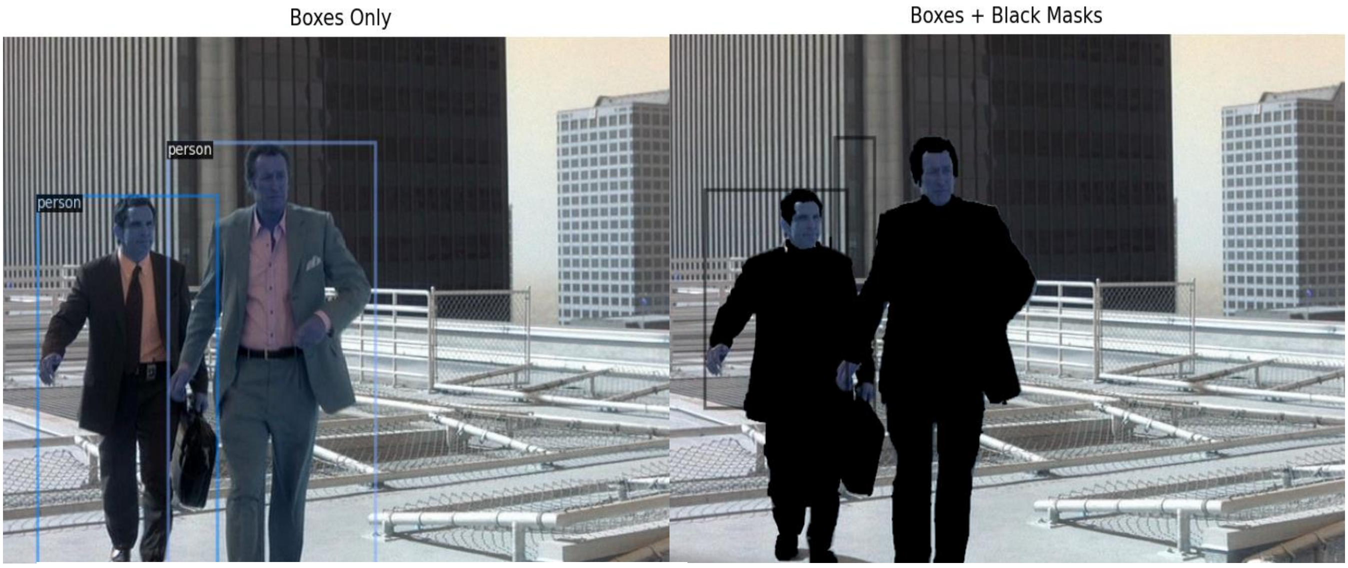
The dataset is analyzed by visualizing sample images with their bounding boxes and parsing masks, and human errors in the annotations are examined. The report then explores how to train a model to detect and segment people in images using this dataset. An attempt is made to build a model capable of classifying the gender of the detected individuals.

The challenges encountered while working with the dataset are discussed, along with the steps taken to address them. Finally, a model is trained to perform both gender classification and instance segmentation, with a focus on visualizing only the masks of female individuals.

Dataset and Tools

The dataset contains a diverse collection of images featuring a wide variety of subjects and scenes. It includes images of children, men, and women engaged in different activities, such as sports, outdoor adventures, indoor tasks, and social gatherings. The dataset captures individuals in various settings, including professional environments, recreational activities, and everyday life scenarios. The subjects represent a broad spectrum of ethnicities, ages, and genders, ensuring a balanced and inclusive representation.

For each image in the dataset, there are two key types of annotations: bounding boxes and masks. Each person in the image is annotated with a bounding box, which is a rectangular region defined by coordinates (x1, y1, x2, y2) that tightly encloses the individual, indicating their presence and location in the image. These bounding boxes are useful for tasks like object detection and localization. Additionally, each person is annotated with a segmentation mask that precisely outlines their body, excluding the face.



Model Train

The next step, after understanding the dataset, is to train a model for detecting people using the LV-MHP-v2 dataset. This model will focus solely on person detection, leveraging the bounding boxes and masks provided in the dataset to localize and segment individuals in the images.

Evaluation:

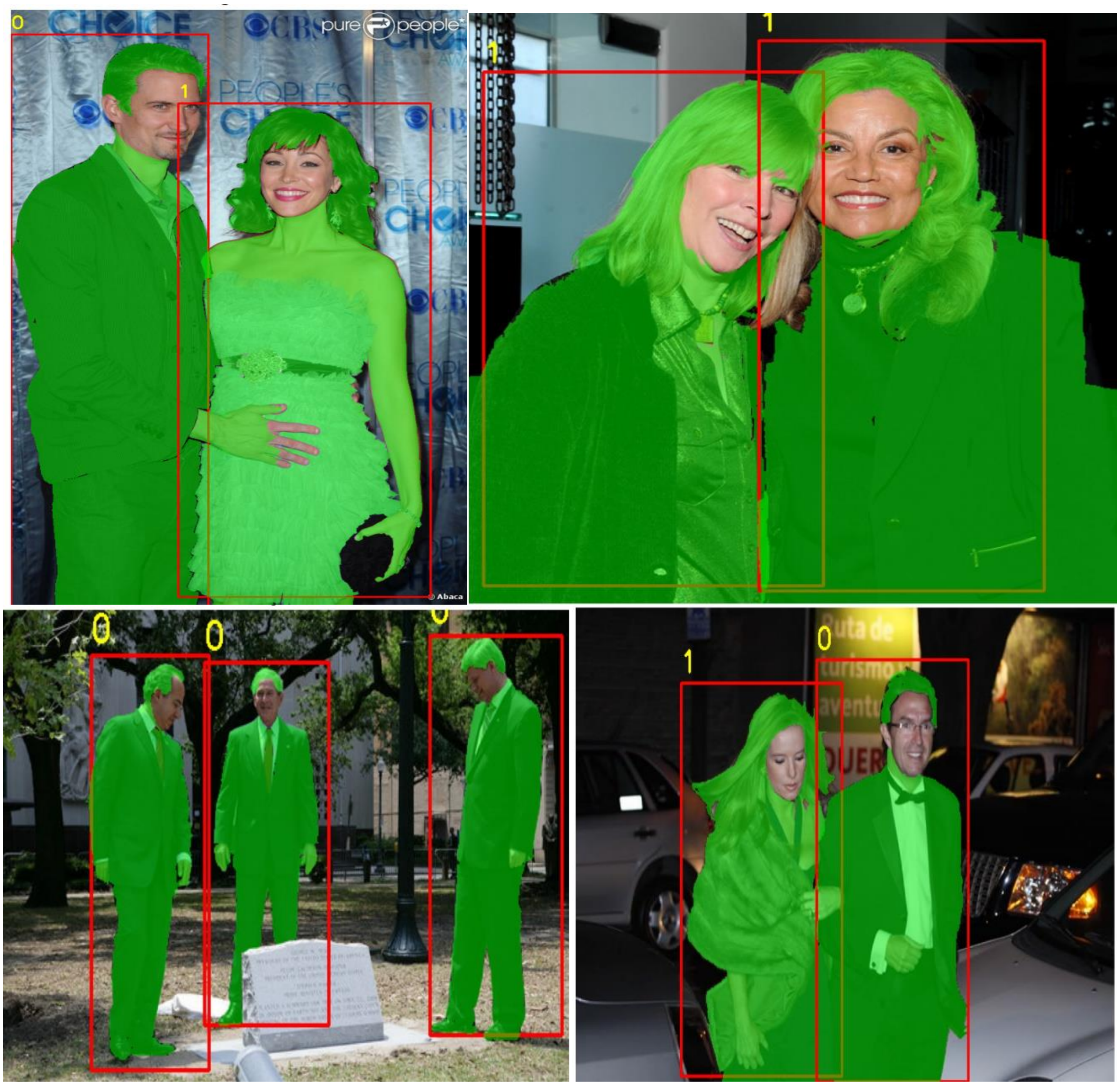
AP	AP50	AP75	APm	APl
66.736	91.026	75.698	51.497	66.846

The next step is to train a model for gender detection. However, the current model lacks gender classification features, which is essential for the task. To address this, additional data will need to be generated to include gender labels. The CVAT (Computer Vision Annotation Tool) website has been chosen as the platform to create this new dataset. Using CVAT, bounding boxes will be annotated for each person in the images, and they will be classified into gender categories (child, male, female). Enabling it to not only detect people but also classify them by gender.

The dataset currently does not include masks, so it can only be used for training a model to detect and classify gender. Without masks, tasks like instance segmentation or detailed body segmentation are not feasible. However, the bounding boxes and gender annotations are sufficient for training a gender detection model.

To better understand the masks, I attempted to train the LV-MHP-v2 dataset for person segmentation. During this process, I discovered that many images in the dataset contained incorrect or inaccurate masks.

Now, to build the final model after understanding all the processes, neither dataset is sufficient on its own. The LV-MHP-v2 dataset lacks gender annotations, and the generated dataset lacks masks. Using two separate models—one for detection and another for masking—is a possible solution, but it would require more computational resources and could lead to errors propagating between the models if one makes a mistake. To address this, the proposed solution is to combine the paths of all masks with their corresponding bounding boxes by generating a new JSON file. This unified file will integrate the necessary annotations, enabling the model to handle both detection and segmentation tasks efficiently.



Class 0 >> Men   Class 2>>women

After Training:

NOTE: The evaluation results provided are based on the training data due to an issue with the validation JSON file. During evaluation, I encountered an error indicating that the length of the classes did not match, even though I verified that everything appeared to be correct. This suggests there may be a problem with the validation JSON file. Due to time constraints, I was unable to resolve this issue and instead evaluated the model using the training data. I acknowledge that these results are not truly representative, as evaluating on the training data can lead to artificially high performance—if the model learns something incorrectly during training, the evaluation on the same data will still show correct results. However, the images shown above are from the validation set, which the model did not train on, providing a more accurate representation of its performance.

Box:

AP	AP50	AP75	APm	APl
82.4228	99.31	98.49075	86.0291	82.255

Seg

AP	AP50	AP75	APm	APl
56.385	90.476	65.569	45.251	57.441

Average Precision (AP) @[IoU=0.50:0.95 | area= all | maxDets=100] = 0.564

Average Precision (AP) @[IoU=0.50 | area= all | maxDets=100] = 0.905

Average Precision (AP) @[IoU=0.75 | area= all | maxDets=100] = 0.656

Average Precision (AP) @[IoU=0.50:0.95 | area= small | maxDets=100] = -1.000

Average Precision (AP) @[IoU=0.50:0.95 | area=medium | maxDets=100] = 0.453

Average Precision (AP) @[IoU=0.50:0.95 | area= large | maxDets=100] = 0.574

Average Recall (AR) @[IoU=0.50:0.95 | area= all | maxDets= 1] = 0.407

Average Recall (AR) @[IoU=0.50:0.95 | area= all | maxDets= 10] = 0.642

Average Recall (AR) @[IoU=0.50:0.95 | area= all | maxDets=100] = 0.644

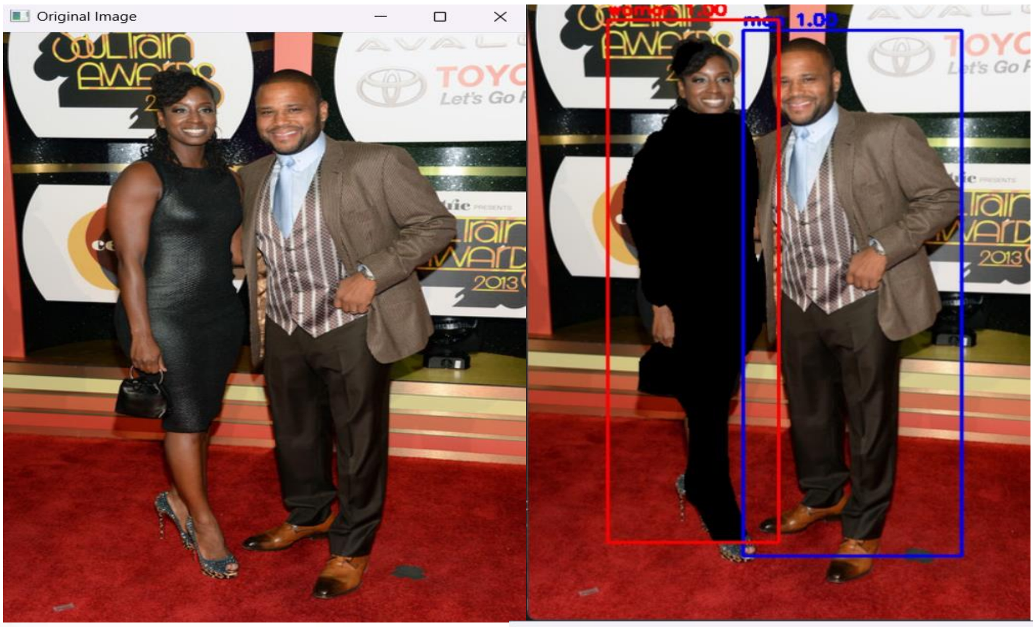
Average Recall (AR) @[IoU=0.50:0.95 | area= small | maxDets=100] = -1.000

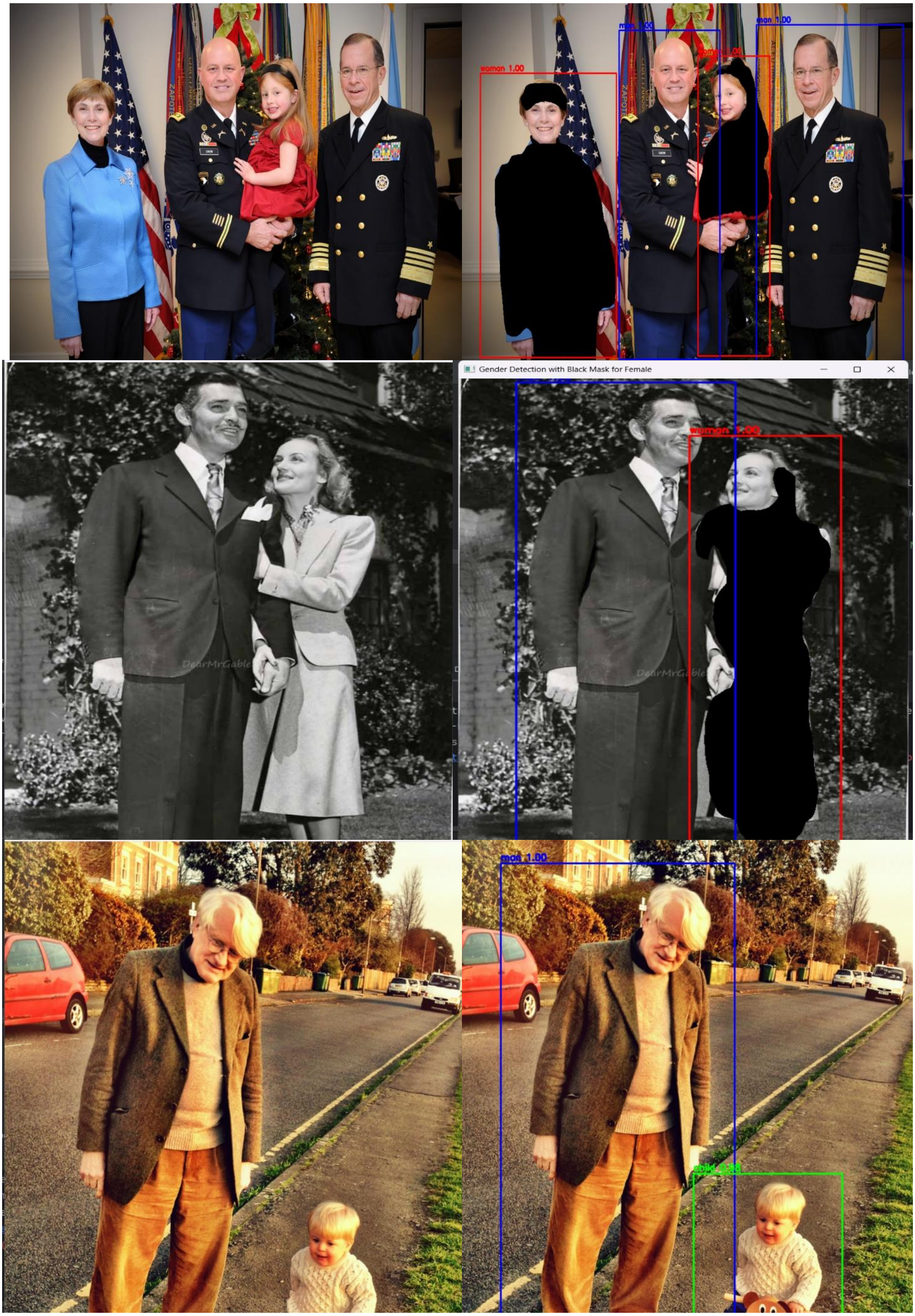
Average Recall (AR) @[IoU=0.50:0.95 | area=medium | maxDets=100] = 0.560

Average Recall (AR) @[IoU=0.50:0.95 | area= large | maxDets=100] = 0.647

Now, to get to the point: we can use the model to detect gender and then retain only the masks for females. This approach allows us to filter and focus on the specific segment of interest—female individuals—while discarding the masks for males and other genders. This step can be particularly useful for applications where only female segmentation is required.

Here are the results obtained by evaluating the model on a random sample of images from the original validation folder. These images were not used during training, ensuring an unbiased assessment of the model's performance. The results demonstrate the model's ability to generalize to unseen data and provide insights into its accuracy and robustness.




woman 1.00
man 1.00
man 1.00
Gender Detection with Black Mask for Female
woman 1.00
man 1.00

Summary of Files and Their Content:

File Name	Content Description
person_detect.py	Code to train a model for detecting persons.
gender_detect.py	Code to train a model for detecting gender (male, female).
segmentation.py	Code to train a model for masking people (body only, excluding faces).
gender_masking_detection.py	Code to train a model that detects gender and generates masks simultaneously.
final_result.py	Code to use the trained model to detect gender and retain only female masks.
output/	Directory containing all trained models and evaluation results.
display_gender_mask.py	Directory containing all trained models and evaluation results.
gender_display.py	Code to display the dataset with bounding boxes only (no masks).
LV-MHP-v2_display.py	Code to display the LV-MHP-v2 dataset with both bounding boxes and masks.
Mask_Gender.json	JSON file containing paths to images and their corresponding masks with gender labels.
coco_cvat.json	JSON file in COCO format containing bounding boxes for gender detection (no masks).