Assignment2

Computer Vision Task

Understanding and Utilizing the Detectron2 Framework for Object Detection Tasks

in this task the objective is to understand LV-MHP-V2 dataset and directories structure, and visualize dataset.

Data set structure:

LV-MHP-v2

- Train
  - Images
  - Parsing annotations
  - Pose annotations
- Val
  - Images
  - Parsing annotations
  - Pose annotations
- Test
  - Images

- **Images:** The dataset consists of JPEG images. Each image is named with an ID number in the format id.jpg, where id ranges from 1 to 25794. The images represent a variety of scenes, each with one or more individuals.

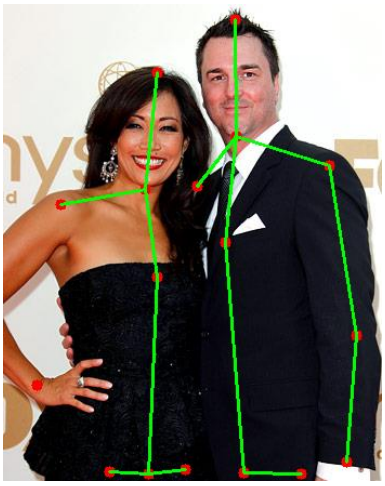
- **Parsing Annotations:** The parsing annotations are provided in PNG format. For each image in the train and val folders, there is a corresponding parsing annotation image. The number of parsing annotation images depends on the number of persons in the corresponding image.
  For an image with ID Id.jpg in the images folder, if the image contains N persons, there will be multiple corresponding parsing annotation files named as Id_02_i.png, where i ranges from 1 to N. Each PNG image contains segmentation masks, where different body parts and clothing items are segmented and assigned a unique color.

- **Pose Annotations**: The pose annotations are stored in MATLAB files. Each MATLAB file corresponds to a specific image and contains the location of keypoints for various body parts. These keypoints are identified by a specific index for each body part in the file.

Annotations Visualization:

(a) Bounding boxes

(b) Parsing annotations



(b) Pose annotations

Figure1: MHP-V2 dataset annotations.

Training Model:

The training done on the given model:

Model: Mask R-CNN with a ResNet-50 backbone and Feature Pyramid Networks (FPN) for instance segmentation.

Hyperparameters:

- Region of Interest (ROI) heads per image: 128
- Number of images per batch: 2
- Learning rate: 0.00025

Maximum number of iterations: 3000

Given the large size of the dataset, even the less accurate ResNet-50 model in Detectron2's segmentation models produces good results.

These results of model's performance to detect persons based on the bounding box predictions at different levels of Intersection over Union (IoU) thresholds:

- AP (Average Precision): 63.48%
- AP at IoU 50% (AP50): 88.80%
- AP at IoU 75% (AP75): 71.80%

Samples from model's predictions:

(a) First sample (b) Second sample

Figure 2: Person Detection Model - Sample Results.

Instance Segmentation and Gender Classification Task

In this task, we develop a model for gender classification and instance segmentation using the MHP-V2 dataset. The objective is:

1- **Gender Classification**: The model classifies each detected person as either male or female.
2- **Selective Instance Segmentation**: For individuals classified as female, the model applies segmentation masks to cover their entire body except for the face and hands.

1. Data Annotation

The MHP-V2 dataset provides annotations solely for the "person" class. For the given task, additional annotations are manually created using the MakeSense website. The dataset is annotated with the following four categories:

- Man (Category ID: 1)
- Woman (Category ID: 2)
- Woman's face (Category ID: 3)
- Woman's hands (Category ID: 4)

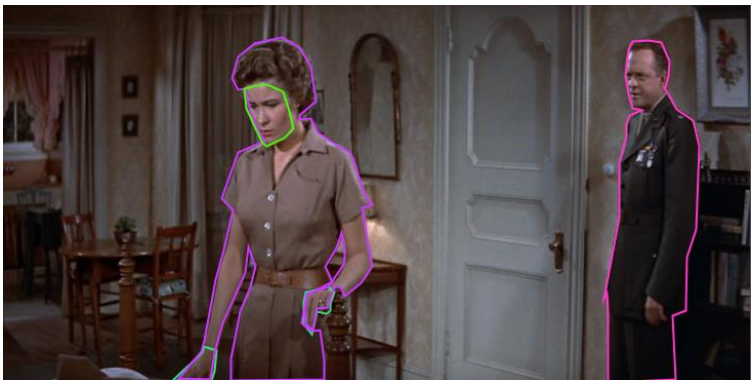


Figure 3: Segmentation process.

These annotations are then exported as a single JSON file in COCO format.

For the training and validation process, the dataset is split as follows:

- The first 300 images (based on image IDs) from the MHP-V2 dataset are used for training.
- The first 100 images from the validation set are used for validation.

2. Model Training

First Model:

Model Architecture: Mask R-CNN with a ResNet-50 backbone and Feature Pyramid Networks (FPN) for instance segmentation.

Hyperparameters:

- Region of Interest (ROI) heads per image: 128
- Number of images per batch: 2
- Learning rate: 0.00025
- Maximum number of iterations: 6000

Performance on Validation set:

Metric	Bounding Box	Mask
Average Precision (AP)	60.06%	57.49%
AP at IoU=0.5 (AP50)	79.21%	79.72%
AP at IoU=0.75 (AP75)	65.48%	66.74%

Table 1: Instance Segmentation first model overall performance metrics.

Category	AP (bbox)	AP (segm)
Men	80.62%	74.27%
Women	75.41%	70.19%
Women's Face	57.77%	59.72%
Women's Hand	26.46%	25.78%

Table 2: Instance Segmentation performance metrics for the first model per category.

Women's hands are poorly detected, but other objects show acceptable results.

Results:

(a) First sample (b) Second sample

Figure 4: Instance Segmentation first model predictions - Sample Results.

(a) First sample (b) Second sample

Figure 5: Instance Segmentation first model predictions with overlayed masks - Sample Results.

Second Model:

Model Architecture: Mask R-CNN with a ResNet-101 backbone and Feature Pyramid Networks (FPN) for instance segmentation (much accurate than ResNet-50 model).

Hyperparameters:

- Region of Interest (ROI) heads per image: 128
- Number of images per batch: 2
- Learning rate: 0.00025
- Maximum number of iterations: 6000

Performance on Validation set:

Metric	Bounding Box	Mask
Average Precision (AP)	61.15%	58.76%
AP at IoU=0.5 (AP50)	81.06%	82.21%
AP at IoU=0.75 (AP75)	68.96%	66.54%

Table 3: Instance Segmentation second model overall performance metrics.

Category	AP (bbox)	AP (segm)
Men	80%	75.51%
Women	80.32%	73.86%
Women's Face	57.72%	59.30%
Women's Hand	26.55%	26.38%

Table 4: Instance Segmentation performance metrics for the second model per category.

Results:

(a) First sample (b) Second sample

Figure 6: Instance Segmentation second model predictions - Sample Results.

(a) First sample (b) Second sample

Figure 7: Instance Segmentation second model predictions with overlayed masks - Sample Results.