| INTRODUCTION TO COMPUTER VISION: VISUAL PERCEPTION AND GENERATION | | |
|---|---|---|
| Student's Name | AIMS African Institute for Mathematical Sciences SENEGAL | Deadline |
| **Uriel Nguefack Yefou** | | **16.05.23, 23:59 pm** |
| May 16, 2023 | | Ac. Year: 2022 - 2023 |
| Lecturer(s): "**Natalia Neverova**" | | |

**LAB1: Report**

# 1 Part B: Instance Segmentation

1. **Model Architecture:** The model we used for instance segmentation was Mask R-CNN with a ResNet-50-FPN backbone pre-trained on the COCO dataset. Mask R-CNN is a two-stage framework that first generates object proposals and then refines them into final detections using a mask head. The COCO dataset is a large-scale dataset of approximatively 120k images annotated with bounding boxes and object segmentation masks for 80 object categories, and keypoint locations for 17 keypoints on the human body.

2. **Examples:**

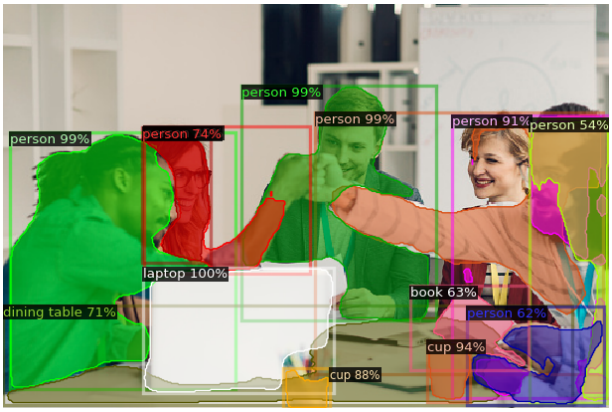


Figure 1: Correct Prediction 1: The model correctly identifies the persons and other objects(book, laptop,..) and segments them.

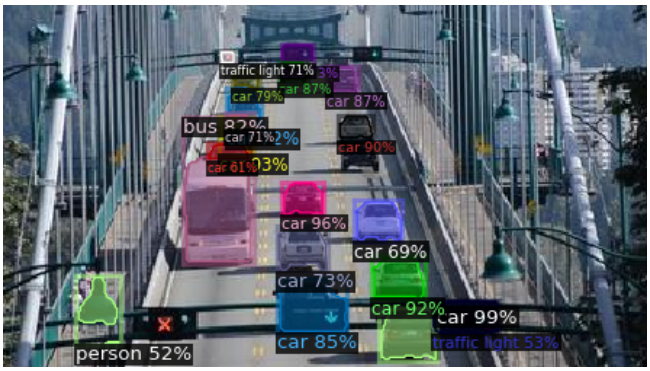


Figure 2: Correct Prediction 2: The model correctly identifies the persons and cars and segments them.


Figure 3: Incorrect Prediction 1: The model fails to detect the eye glass and the book.


Figure 4: Incorrect Prediction 2: The model incorrectly detects the passion fruits and segments them as if there were oranges and apples.

3. **Observations:**

We observed that the model performs well in images with clear and unobstructed object instances. However, it struggles with occluded and small object instances, often failing to detect them. Another obser-

vation is that the model seems to be highly sensitive to the color and texture of objects, segmenting them as other objects.

4. **Error Modes:** We discovered that the model tends to fail in detecting small object instances, often missing them altogether. Additionally, the model seems to be confused by object instances that are partially occluded, segmenting the visible parts and leaving out the occluded parts. We also observed that the model makes more errors when there are multiple object instances in the image, as it may confuse them with one another.

## 2 Part C: Human Pose Estimation

1. **Model Architecture:** The model we used for human pose estimation was Keypoint R-CNN with a ResNet-50-FPN backbone pre-trained on the COCO dataset. Keypoint R-CNN is a two-stage framework that generates object proposals and then predicts keypoint locations within each proposal. The data is still the COCO dataset.

2. **Examples:**


Figure 5: Correct Prediction 1: The model correctly identifies the persons and detects the keypoint locations on their bodies.

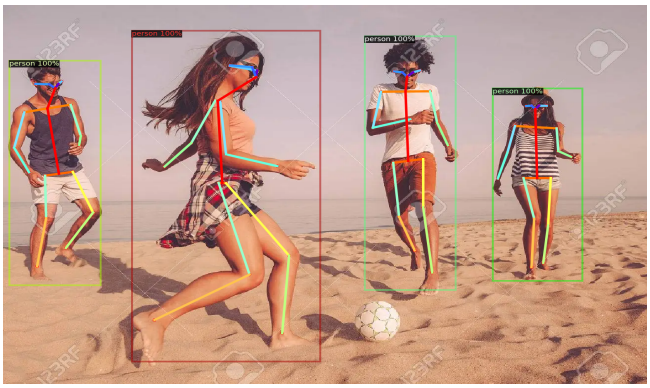


Figure 6: Correct Prediction 2: The model correctly identifies the persons and detects the keypoint locations on their bodies.

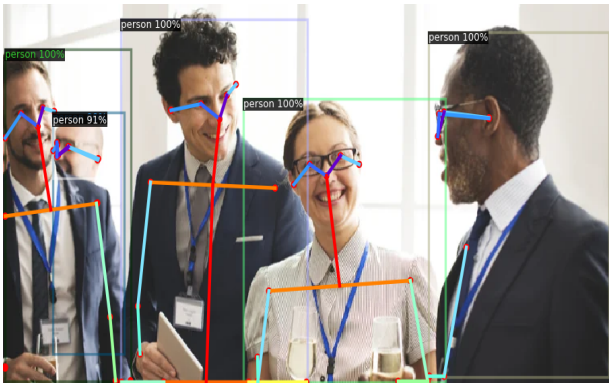


Figure 7: Incorrect Prediction 1: The model fails to detect some keypoints locations on the person's body.

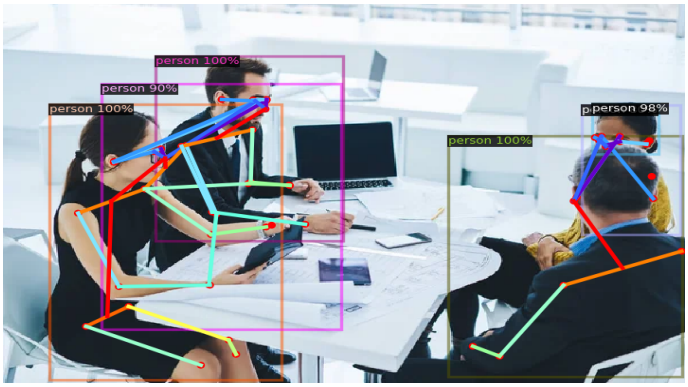


Figure 8: Incorrect Prediction 2: The model incorrectly detects some keypoints and make multiple prediction on same image.

3. **Observations:**

   We observed that the model performs well in images with clear and unobstructed bodies parts. However, it struggles with occluded part of the body, often failing to detect them. Another observation is that the model seems to be highly sensitive when people are too close together, in this case it puts too much keypoints.

4. **Error Modes:** We discovered that the model tends to fail in detecting keypoints in occluded parts of the body. We also observed that the model makes more errors when there are people too close together.