

# LAB1: Detectron2 Warmup

## Computer Vision 2

Mory Moussou Koulibaly Traore

AMMI/AIMS April 2021

## 1 INTRODUCTION

The purpose of this lab was to familiarize ourselves with Detection2, library that implements state-of-the art object detection models. It can load models such as Mask R CNN, Faster R CNN. For Mask R CNN we can have for example ResNet-50-FPN as a backbone. And the pre-trained model can be run for instance segmentation, human pose segmentation and panoptic segmentation (combines semantic and instance segmentation). Now let us summarize your findings and exploration with Detectron2.

## 2 Part B:Run a pre-trained model for instance segmentation

### 2.1 Model architecture

Here the task is to run a pre-trained model for instance segmentation. This is for showing for each image the instance of every Known object. For this we have loaded the Mask R CNN pre-trained model from Detectron2 Zoo library. Mask R CNN use here the Resnet-50-FPN backbone. Mask R-CNN adopts the same two-stage procedure, with an identical first stage (which is RPN). In the second stage, in parallel to predicting the class and box offset, Mask R-CNN also outputs a binary mask for each ROI. The models were pre-trained on COCO dataset. COCO is a dataset of 120k images annotated with bounding boxes and object segmentation masks for 80 object categories.

### 2.2 Example: Visualisations instance segmentation

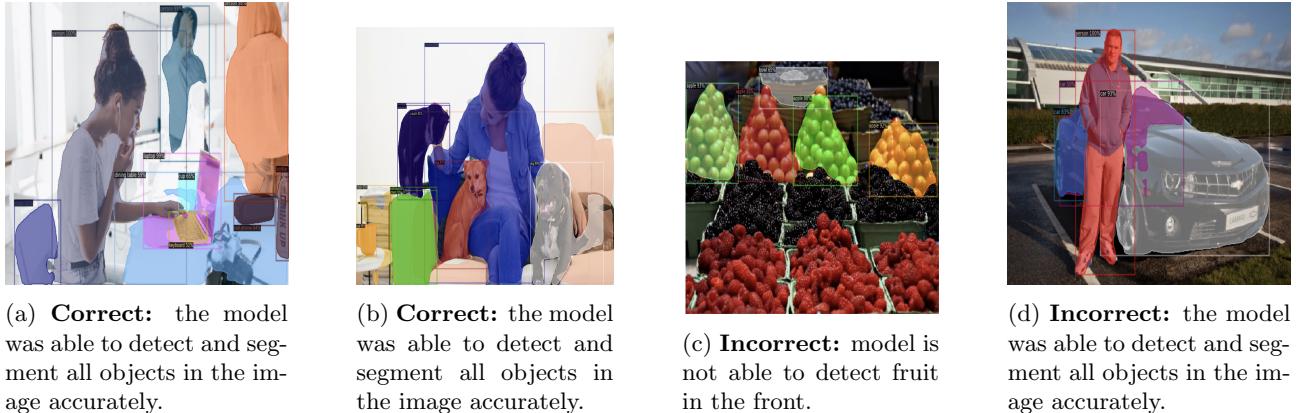


Figure 1: Visualisations Instance Segmentation

### 2.3 Observations/error

For this task of instance segmentation, we realize that the models not able to capture objects when they are very far and also when we have too many objects in the image which are almost overlapping. The error is We see this this behavior also in different images.

### 3 Part C: Run a pre-trained model for pose estimation

#### 3.1 3.1 Model Architecture

Part C concerns human pose estimation which is a task that consists of detecting people and localizing their keypoints detection within an image such as eyes, nose, shoulders. The model loaded is also R-CNN . The goal of R-CNN is to take in an image, and correctly identify where the primary objects (via a bounding box) in the picture. Inputs: Image; Outputs: Bounding boxes and labels for every object in images. And R-CNN is just the following steps: generate a set of region proposals for bounding boxes, run the images in the bounding boxes through a pre-trained AlexNet( Resnet-50-FPN here) and finally an SVM to see what object the image in the box is, run the box through a linear regression model to output tighter coordinates for the box once the object has been classified.

#### 3.2 Example: Visualisations of predictions human pose estimation

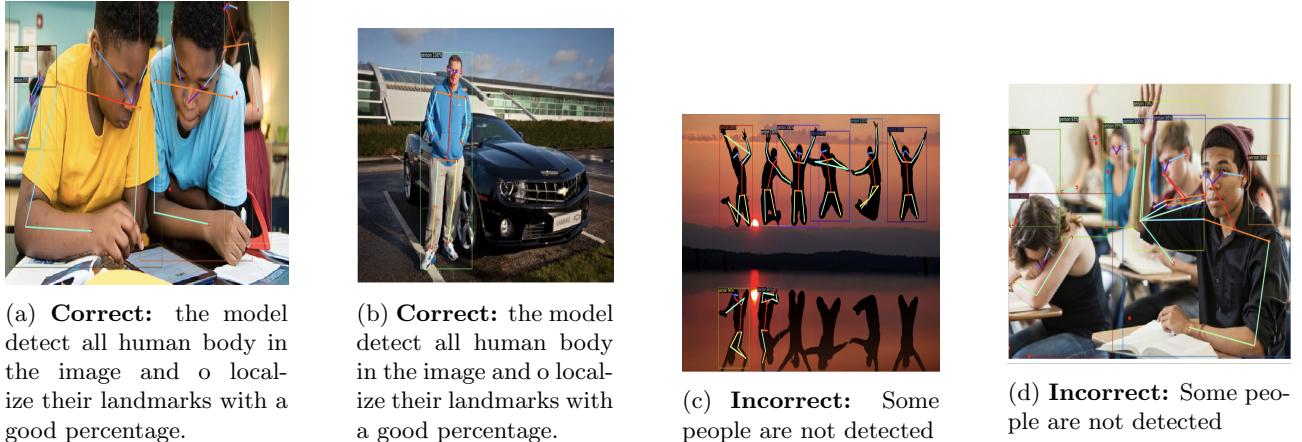


Figure 2: Visualisations Instance Segmentation

#### 3.3 Observations/error

For keypoints detection, sometimes the model can't detect people and give landmarks of images when they are also far away in the image. The error is the model were able to work well for many examples even if the quality is not so good.

## 4 Conclusion

In this lab1, we learned many thing for using detectron2. We know the installation, how to use pre-trained models in different tasks of object detection. And with our observations We see that the models have can predict in most cases but make some mistakes in others