AFRICAN MASTERS IN MACHINE INTELLIGENCE

(AMMI RWANDA, KIGALI)

Name: Aimable ISHIMWE MANZI
Course: Computer Vision 2

Lab Number: 1
Date: May 6, 2020

REPORT OF LAB1 ON DETECTRON2

1. **Introduction**

   In this practitical session of object detection library implemented using PyTorch, we learn and implement object detection models using Detectron 2. It provides a model zoo, a library of pre-trained models trained on a variety of datasets. A model architecture consiste of COCO dataset over 300,000 images which are annotated with bounding boxes and instance masks of 80 object categories, was run on an image to perform this task of instance segmentation and object detection [1].

2. **Pretrained instance segmentation**

   In this part we load an Image and use a pre-trained model. The datasets used helped as on understanding of visual scene. Objects are labeled using per-instance segmentations to aid in precise object localization.[3] This example tutorial in Detectron2 shows a model with ResNet50 backbone pretrained for instance segmentation and object detection tasks in the COCO dataset. Below is the image showing pre-trained model applied on image and able to predict the content of image separately with it's corresponding accuracy.

Figure 1: draw the predictions on the image

3. **Pretrained pose estimation**

In this part we continued using dectron2 strucure in bounding box. The task was to use pre-trained models to process the task of human pose estimation, then after we visualise the predictions. The architecture consist of Mask R-CNN (faster R-CNN + keypoint estimation) model along with a ResNet50 backbone architecture which was pretrained on the COCO dataset was run to perform human pose estimation task.

(a) Inference with a keypoint detection model

(b) Inference with a panoptic segmentation model

Figure 2: plots of different builtin models

4. **Conclusion or Evaluation**

- By visualizing prediction on Image 1, the prediction on a person and laptop have accuracy more than 80%. This shows that the pre-trained model managed to identify object and person. On the other part of pose estimation, It perfectly managed to view all k-points as seen in Image 2a, all person's eye are pointed and all joint are pointed. The same to image 2b, it managed to mask all object and person found.
- Correct prediction: we choose the following correct preduction basing on the accuracy of how the model identified it compaired to the other objects.
  (a) It manage to predict a person in image 1.
  (b) It predict bottle, laptop and chair more that 70% in image 1.
  (c) Perfectly masking person, laptop, banner, and bottles with accuracy over 90% in Image 2b.
- False Prediction: The identified false preduction were chosen based on our own visualisation and accuracy below 70%.
  (a) Bag in image 1.
  (b) One lady predict as 56% in image 1.
  (c) In image 2b it mask one part of a wall as window.
- The errors in models that I have seen keep coming are:
  (a) Clouded picture, this comes as an error, since the model isn't able to detect all objects that are in the picture.
  (b) Images that are close(not far): I found out that this comes as an error as the model doesn't predict it well.
  (c) My insight on what causing errors is that, some object which are not in the model and even one in model, it give small accuracy or ignore.

References

[1] https://colab.research.google.com/drive/1Qs6m-gZ7It53hmMbCNGST962cycQWRvW

[2] https://github.com/gkioxari/aims2020-visualrecognition/blob/master/LAB1.md

[3] *Microsoft coco: Common objects in context*, Lin, Tsung-Yi and Maire, Michael and Belongie, Serge and Hays, James and Perona, Pietro and Ramanan, Deva and Dollár, Piotr and Zitnick, C Lawrence, European conference on computer vision, 2015.