Homework 12

Semantic Image Segmentation

Description

In this assignment, you will apply the DeepLabV3 and Lite R-ASPP models to perform semantic segmentation in videos.

Tasks

The specific steps for this task are:

Part A:

1. Are image segmentation models considered multi-task architectures, and what does "multi-task" mean in this context?

Part B:

1. Use a video file of your choice to perform this assignment. The video should be low-resolution and short (i.e. 5-10 seconds, 240x480 pixels). Please pay attention to the licensing and references when using online video repositories.
2. Write a Python code using OpenCV to parse video frames and perform semantic segmentation of the video using:
3. DeepLabV3 architecture with MobileNetV3 weights [2].
4. Lite R-ASPP architecture with MobileNetV3 weights [2].
5. Which model seems to perform better ? (since you do not have groundtruth you will have to perform visual inspection for a qualitative comparison and provide your comments)
6. What is the main difference DeepLabV3 and LR-ASPP? Please elaborate.

Submission Guidelines

- Submit your working code in Teams (**both as an .ipynb and a .pdf file**)
- Upload any .zip file or folder if your code refers to the paths of those files.
- **A pdf of your report (name: HW13- PartX-Report-Firstname-Lastname.pdf) with your output and comments**

References

[1] “Semantic Segmentation using PyTorch”, https://debuggercafe.com/semantic-segmentation-using-pytorch-deeplabv3-and-lite-r-aspp-with-mobilenetv3-backbone/

[2] “MODELS AND PRE-TRAINED WEIGHTS”, Pytorch https://pytorch.org/vision/stable/models.html

[3] “Getting Started with Videos”, OpenCV
https://docs.opencv.org/3.4/dd/d43/tutorial_py_video_display.html