Progress Update: Motion Segmentation

12/01/2022

Introduction

In this report we summarise the work that was produced over the Christmas break from 22/12/2021 - 12/01/2022.

Progress Update

Training runs

We extended our Tensorboard tracking to also plot IoU after each epoch. Below we show the best curve for aIoU on the validation set.

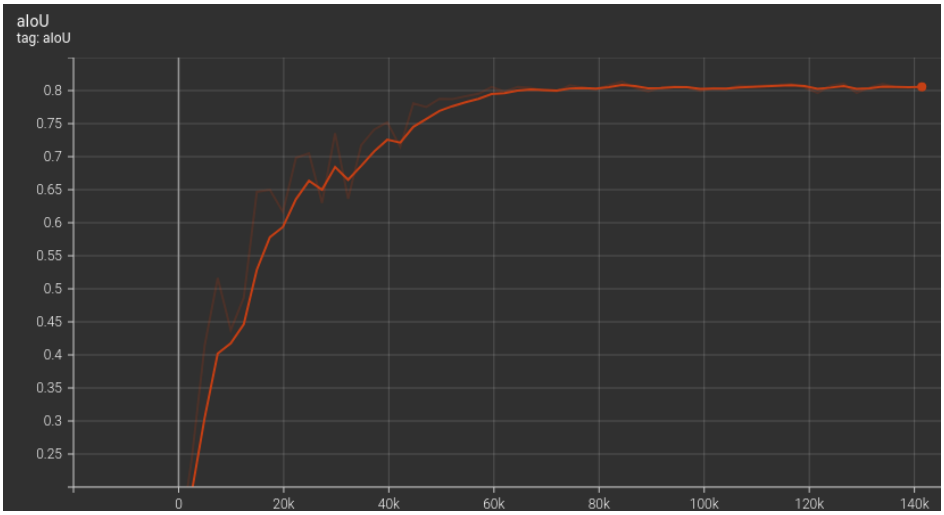

Figure 1: Best IoU achieved on Extended Kitti Mod

We held back one sequence to perform testing on. The IoU on the test was 0.23. This large discrepancy between validation and test is most likely due to only using one sequence as test which might not represent the rest of the data well.

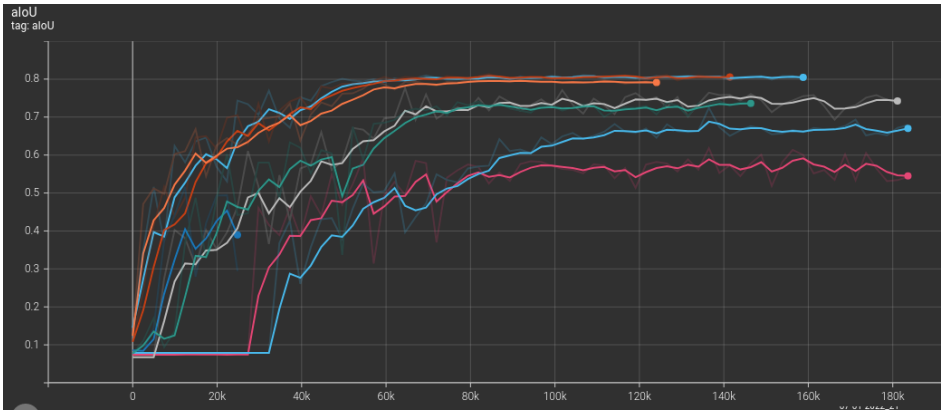

Figure 2: Different training runs

Focal loss parameters

Default from the paper [1] $\gamma = 2.0$ and $\alpha = 0.25$

Datasets

- Extended KITTI ($\sim$ 1200 images)
- Extended KITTI Mod ($\sim$ 12,000 images of which 4000 are usable)
- Video-agnostic KITTI (COCO annotations separated by only static/only dynamic/mixed)

For *Video-agnostic KITTI* we cannot filter for cars only/pedestrians only. Therefore we cannot compare it to the *Extended KITTI Mod*. Also there are only about 2500 images with moving objects in the *Video-agnostic KITTI* dataset which is not enough to get the model to generalise well.

Either use real data from KITTI, Cityscapes, etc (just one of them) or create our own dataset from CARLA. We will have more control, can collect more data but the data is artificially generated.

CARLA

Static vehicles as actors

When spawning static vehicles, semantic segmentation will show the statically set actors as vehicles as well and it will be difficult to generate the motion maps.

Semantic segmentation: Obtain instances with semantic segmentation and depth map

Instance segmentation (from CARLA 09.13 - available locally): Use instance segmentation together with optical flow maps to differentiate between static and dynamic actors and generate motion map.

Static vehicles embedded in the map

Using Matlab's Roadrunner, we can modify the map to include more static vehicles and at other locations. By having a separate semantic tag it will be easier to generate the ground truth of dynamic pixels (spawned moving actors) and static pixels (ParkedVehicles map incorporating stationary vehicles)

Next Steps

- Carla spawn static vehicles $\rightarrow$ Use Roadrunner?
- Different network architectures $\rightarrow$ Siamese network/completely different architecture?
- Other potential data sources $\rightarrow$ not likely to exist
- Unsupervised extension $\rightarrow$ how to proceed with supervised to make it unsupervised

References

[1] T. Lin, P. Goyal, R. B. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in *IEEE International Conference on Computer Vision, ICCV 2017, Venice, Italy, October 22-29, 2017*. IEEE Computer Society, 2017, pp. 2999–3007. [Online]. Available: https://doi.org/10.1109/ICCV.2017.324