

Progress Update: Motion Segmentation

02/02/2022

Introduction

We have continued to resolve our data problems. In regards to the real data, we seem to have identified the problem which was producing a large validation-test performance gap. We changed the way we split datasets and how training data is produced in the *Dataset* class. Additionally, we also performed multiple grid searches on this changed data and datasets to optimize parameters.

Carla's Town03 was modified with around 150 vehicles parked (static) cars. The ground truth is obtained by using the semantic segmentation and lifting pixels having a walker or dynamic vehicles semantic label only. Additionally, each dynamic actor is checked whether it is actually moving or not (red light, traffic jam etc).

KITTI Mod Seg Dataset

Changes

We have changed the way we split our datasets since the discussion in last week's meeting about how to perform dataset-splitting when the data consists of sequences. We now don't hold back entire sequences to perform testing on. Instead, we shuffle and split the entire dataset into train, validation and test before we start training. This leads to the test set having a much more similar distribution to the other two sets. This closes the validation-test performance gap we experienced previously.

The second change we made was that we no longer have overlapping images between image pairs coming from the dataloader. Meaning that there now is no possibility of any of the datasets containing the same image. This change also helps close the validation-test performance gap.

Finally, we performed multiple grid searches using a subset of our new datasets to find the best *learning rate*, *patience* and *learning rate scheduler factor*. Using the best parameters we are now currently training a model on the entire dataset.

Observations

Below we discuss two things we noticed about the data and our model's predictions in the last week.



Figure 1: Problematic Ground Truth Example

One factor which most likely still contributes to bad performance is the ground truth data. As can be seen in Figure 1, the car in the left image is labeled as a rectangle. In this case our model actually made better predictions than the ground truth but the IoU metric which we use to evaluate performance penalizes this.

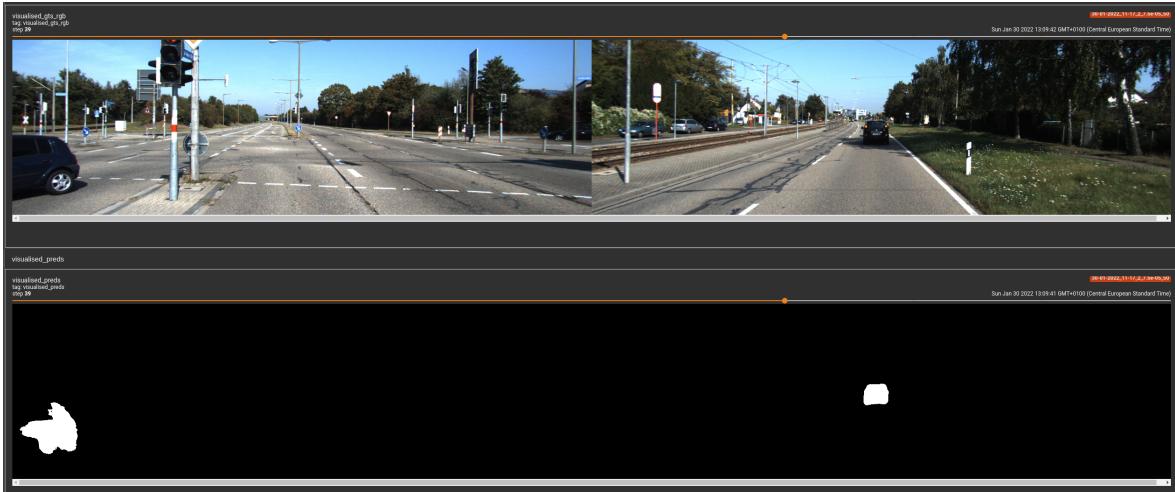


Figure 2: Common Issue with Model Predictions

Figure 2 shows the RGB images at the top and the predictions by our model at the bottom. The predictions for the car on the left is very poor, which was observed to be a common scenario for other images as well. We suspect this has to do with the fact that most images show cars head-on/from behind and proportionally few images show cars passing the ego vehicle. Additionally, most images in the dataset show cars in the distance, so perhaps the model also struggles to detect close-up cars which appear larger. A possible improvement would be to upsample such images to help the model achieve better results.

CARLA Data Generation

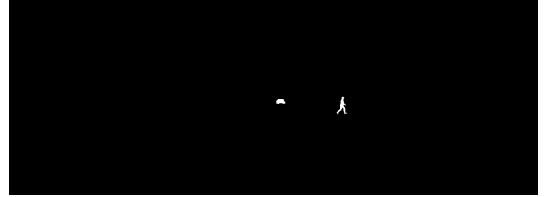
Previously the ground truth for motion segmentation was obtained from semantic labels for parked vehicles, manually placed on the map, and dynamic actors. However, we cannot separate our data in this way since cars standing on red lights or elsewhere would still be marked as dynamic. Instead, now we obtain the instance IDs of all the spawned vehicles

and check if the actor's speed is below a certain threshold. In this way we can now properly segment the moving and non moving pixels.

- 8500 CARLA images without pedestrian annotations
- 3000+ CARLA images with pedestrian annotation & growing
- Added around 150 stationary vehicles to Town03_Opt with additional semantic label (red in the CityScapesColorPalette used for Semantic Segmentation)
- Included motion segmentation for pedestrians
- Randomize weather and ego vehicle spawning
- Changed motion segmentation to only save actual moving cars and not all the spawned moving vehicles. In the example below: The car ahead of us is waiting at the red light. The motion segmentation ground truth only includes the vehicle moving in the tunnel and the pedestrian crossing the road, see Figure 4.



(a) RGB - Image 01



(b) Motionseg - Image 01

Figure 3: Input data vs ground truth



Figure 4: Semseg - Image 01

Next Steps

- Randomize actor vehicle speeds
- Achieve good performance on KITTI Mod Seg Dataset with new datasets split - results pending
- Hopefully achieve even better results on CARLA data, due to the perfect labels
- Start with the self-supervised architecture