CoVLA: Comprehensive Vision-Language-Action Dataset for Autonomous Driving
— Supplementary Material —

A. Heuristic Trajectory Filtering

In some instances, trajectory data instability was detected. Specifically, we identified two erroneous behaviors:

1. Significant jumps

2. Movement in the wrong direction

To detect significant jumps, we filtered the trajectory data by the distance between adjacent points. Given a recording frequency of 20 Hz and a maximum speed of 100 km/h, the distance between points should be at most 1.38 meters, which is calculated as following:

$$\frac{100\,\text{km/h} \times 1000\,\text{m/km}}{3600\,\text{s/h} \times 20} = 1.38\,\text{m} \tag{1}$$

With a tolerance rate of 1.15, the threshold was set to 1.59 meters. All trajectories exceeding this threshold were filtered.

To detect movement in the wrong direction, we manually checked 400 samples from all scenes and identified 43 invalid trajectories (10.75%). Observations revealed a vibration frequency of 10 Hz in these trajectories. To implement vibration detection, we smoothed the trajectory using a 3-point moving average and calculated the difference between the smoothed trajectory and the original trajectory. We then analyzed the variance of these differences. If the variance exceeded a certain threshold, the trajectory was classified as invalid. This method yielded a precision of 0.64 and a recall of 0.75 on the test dataset, reducing the invalid trajectory rate to 2.6%. Although this method has a relatively high false-positive rate, it is acceptable for the dataset's scale.

B. Privacy Protection for Dataset Publication

We published CoVLA-Dataset [3] on HuggingFace. For privacy protection, we anonymized human faces and license plates in the CoVLA-Dataset images and videos, using Dashcam Anonymizer [4].

C. Example data in CoVLA-Dataset

We present sample data from CoVLA-Dataset. Each data point includes an image, a caption, and vehicle states. An example of an actual caption is shown in Figure 7. For auto-captioning, we use the traffic light detection provided by OpenLenda-s and the front car detection results from the sensor fusion. The list of vehicle states is indicated in Table 4. Each frame contains information on the speed, steering angle, coordinates of the trajectory, and more.

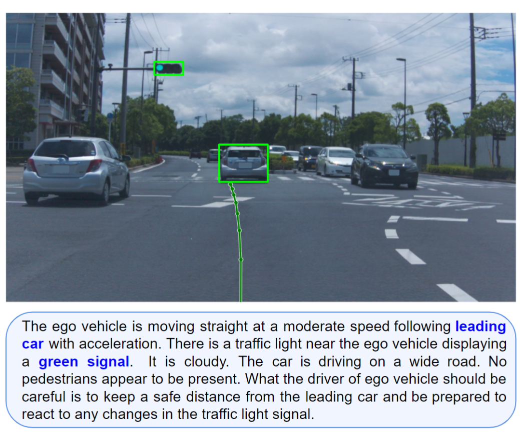

Figure 7. Example of image and caption in CoVLA-Dataset.

[3] https://huggingface.co/datasets/turing-motors/CoVLA-Dataset
[4] https://github.com/varungupta31/dashcam_anonymizer

Key	Value
frame_id	569
image_path	images/2022-07-08–11-37-27–5_first/0569.png
vEgo	7.43082332611084
vEgoRaw	7.4395833015441895
aEgo	0.6044138669967651
steeringAngleDeg	0.6073870658874512
steeringTorque	69.0
brake	0.0
brakePressed	false
gas	0.20499999821186066
gasPressed	true
doorOpen	false
seatbeltUnlatched	false
gearShifter	drive
leftBlinker	false
rightBlinker	false
orientations_calib	$\begin{bmatrix} 2.9467956252753775 & 0.9174552319868815 & 2.2181786819453384 \end{bmatrix}$
orientations_ecef	$\begin{bmatrix} 2.9243210649189977 & 0.9224135550861058 & 2.1900513923432348 \end{bmatrix}$
orientations_ned	$\begin{bmatrix} -0.013463193567253392 & 0.006326533926443111 & -2.990125370637735 \end{bmatrix}$
positions_ecef	$\begin{bmatrix} -3959574.486029379 & 3328427.354910454 & 3719065.7393601397 \end{bmatrix}$
velocities_calib	$\begin{bmatrix} 7.317097759615114 & 0.003242519329502727 & 0.005369323447773883 \end{bmatrix}$
velocities_ecef	$\begin{bmatrix} -2.6767882666706004 & 3.547338396353873 & -5.813015899212604 \end{bmatrix}$
accelerations_calib	$\begin{bmatrix} 0.4734579094803297 & 0.08559864698994124 & -0.13132037594775653 \end{bmatrix}$
accelerations_device	$\begin{bmatrix} 0.4736293760658642 & 0.07819260264673351 & -0.13526157094253702 \end{bmatrix}$
angular_velocities_calib	$\begin{bmatrix} 0.011550541795216845 & 0.012243857869171634 & -0.007753300486330907 \end{bmatrix}$
angular_velocities_device	$\begin{bmatrix} 0.011675888068301523 & 0.01206267096884485 & -0.007848971055415363 \end{bmatrix}$
timestamp	1657248173200
extrinsic_matrix	$\begin{bmatrix} -0.015688330416257182 & -0.9998769191404183 & 0.00012959444326649344 & 0.0 \\ -0.008260370686184616 & 2.879912020664621e-21 & -0.9999658837914467 & 1.2200000286102295 \\ 0.9998428078989188 & -0.01568886620613436 & -0.008259354077745229 & 0.0 \\ 0.0 & 0.0 & 0.0 & 1.0 \end{bmatrix}$
intrinsic_matrix	$\begin{bmatrix} 2648.0 & 0.0 & 964.0 \\ 0.0 & 2648.0 & 604.0 \\ 0.0 & 0.0 & 1.0 \end{bmatrix}$
trajectory_count	60

trajectory	$\begin{bmatrix} 0.0 & -0.0 & 0.0 \\ \dots & \dots & \dots \\ 2.2367286927600376 & 0.021314714278469867 & 0.011369742248091336 \\ \dots & \dots & \dots \\ 4.529136516056675 & 0.05988551136966269 & 0.03572205827208557 \\ \dots & \dots & \dots \\ 6.919731941586808 & 0.08958899605082428 & 0.050459818682157966 \\ \dots & \dots & \dots \\ 9.301852576186251 & 0.1417775525229876 & 0.07520389298492622 \\ \dots & \dots & \dots \\ 11.677091075613516 & 0.2108797042962825 & 0.09371664992653018 \\ \dots & \dots & \dots \\ 14.027571172292527 & 0.28508464282229706 & 0.10760938523286476 \\ \dots & \dots & \dots \\ 16.350994760951718 & 0.38061347529509826 & 0.1208462683905208 \\ \dots & \dots & \dots \\ 18.675065252972097 & 0.4777731491748536 & 0.14622174266379429 \\ \dots & \dots & \dots \\ 20.998125620182435 & 0.5860796022647292 & 0.1657656398148593 \end{bmatrix}$

Table 4. Example of Vehicle States List.