

Added a new "deep lane guidance" module to the Vector Lanes neural network which fuses features extracted from the video streams with coarse map data, i.e. lane counts and lane connectivities. This architecture achieves a 44% lower error rate on lane topology compared to the previous model, enabling smoother control before lanes and their connectivities becomes visually apparent. This provides a way to make every Autopilot drive as good as someone driving their own commute, yet in a sufficiently general way that adapts for road changes.

Improved overall driving smoothness, without sacrificing latency, through better modeling of system and actuation latency in trajectory planning. Trajectory planner now independently accounts for latency from steering commands to actual steering actuation, as well as acceleration and brake commands to actuation. This results in a trajectory that is a more accurate model of how the vehicle would drive. This allows better downstream controller tracking and smoothness while also allowing a more accurate response during harsh maneuvers.

Improved unprotected left turns with more appropriate speed profile when approaching and exiting median crossover regions, in the presence of high speed cross traffic ("Chuck Cook style" unprotected left turns). This was done by allowing optimizable initial jerk, to mimic the harsh pedal press by a human, when required to go in front of high speed objects. Also improved lateral profile approaching such safety regions to allow for better pose that aligns well for exiting the region. Finally, improved interaction with objects that are entering or waiting inside the median crossover region with better modeling of their future intent.

Added control for arbitrary low-speed moving volumes from Occupancy Network. This also enables finer control for more precise object shapes that cannot be easily represented by a cuboid primitive. This required predicting velocity at every 3D voxel. We may now control for slow-moving UFOs.

Upgraded Occupancy Network to use video instead of images from single time step. This temporal context allows the network to be robust to temporary occlusions and enables prediction of occupancy flow. Also, improved ground truth with semantics-driven outlier rejection, hard example mining, and increasing the dataset size by 2.4x.

Upgraded to a new two-stage architecture to produce object kinematics (e.g. velocity, acceleration, yaw rate) where network compute is allocated $O(\text{objects})$ instead of $O(\text{space})$. This improved velocity estimates for far away crossing vehicles by 20%, while using one tenth of the compute.

Increased smoothness for protected right turns by improving the association of traffic lights with slip lanes vs yield signs with slip lanes. This reduces false slowdowns when there are no relevant objects present and also improves yielding position when they are present.

Reduced false slowdowns near crosswalks. This was done with improved understanding of pedestrian and bicyclist intent based on their motion.

Improved geometry error of ego-relevant lanes by 34% and crossing lanes by 21% with a full Vector Lanes neural network update. Information bottlenecks in the network architecture were eliminated by increasing the size of the per-camera feature extractors, video modules, internals of the autoregressive decoder, and by adding a hard attention mechanism which greatly improved the fine position of lanes.

Made speed profile more comfortable when creeping for visibility, to allow for smoother stops when protecting for potentially occluded objects.

Improved recall of animals by 34% by doubling the size of the auto-labeled training set.

Enabled creeping for visibility at any intersection where objects might cross ego's path, regardless of presence of traffic controls.

Improved accuracy of stopping position in critical scenarios with crossing objects, by allowing dynamic resolution in trajectory optimization to focus more on areas where finer control is essential.

Increased recall of forking lanes by 36% by having topological tokens participate in the attention operations of the autoregressive decoder and by increasing the loss applied to fork tokens during training.

Improved velocity error for pedestrians and bicyclists by 17%, especially when ego is making a turn, by improving the onboard trajectory estimation used as input to the neural network.

Improved recall of object detection, eliminating 26% of missing detections for far away crossing vehicles by tuning the loss function used during training and improving label quality.

Improved object future path prediction in scenarios with high yaw rate by incorporating yaw rate and lateral motion into the likelihood estimation. This helps with objects turning into or away from ego's lane, especially in intersections or cut-in scenarios.

Improved speed when entering highway by better handling of upcoming map speed changes, which increases the confidence of merging onto the highway.

Reduced latency when starting from a stop by accounting for lead vehicle jerk.

Enabled faster identification of red light runners by evaluating their current kinematic state against their expected braking profile.