Comparison among End-to-End Self Driving Solutions using Bench2Drive

Conghao Jin, Tianyi Cao, and Jackson Chen
University of Michigan, Ann Arbor, US

Motivation

- End-to-End self driving technique is crucial for advancing autonomous vehicle technology by simplifying development, enhancing adaptability, and leveraging the power of AI. Under this system, it will be more flexible to do processing towards raw sensor data directly, without predefined rule or decision-making strategies.
- Three of the most popular E2E systems – TCP, UniAD and VAD are critically reviewed and compared. Analyzing the pros and cons of these cutting edge strategies helps build a better understanding of the orientation of current technologies and a more valuable outlook for the future.

Methodology

TCP - Trajectory-guided Control Prediction

Output prediction of E2E autonomous driving:

1. **Trajectory / waypoints** - sophisticate control algorithms → struggle for big turns
2. **Direct control actions** - focus on current state → reaction latency

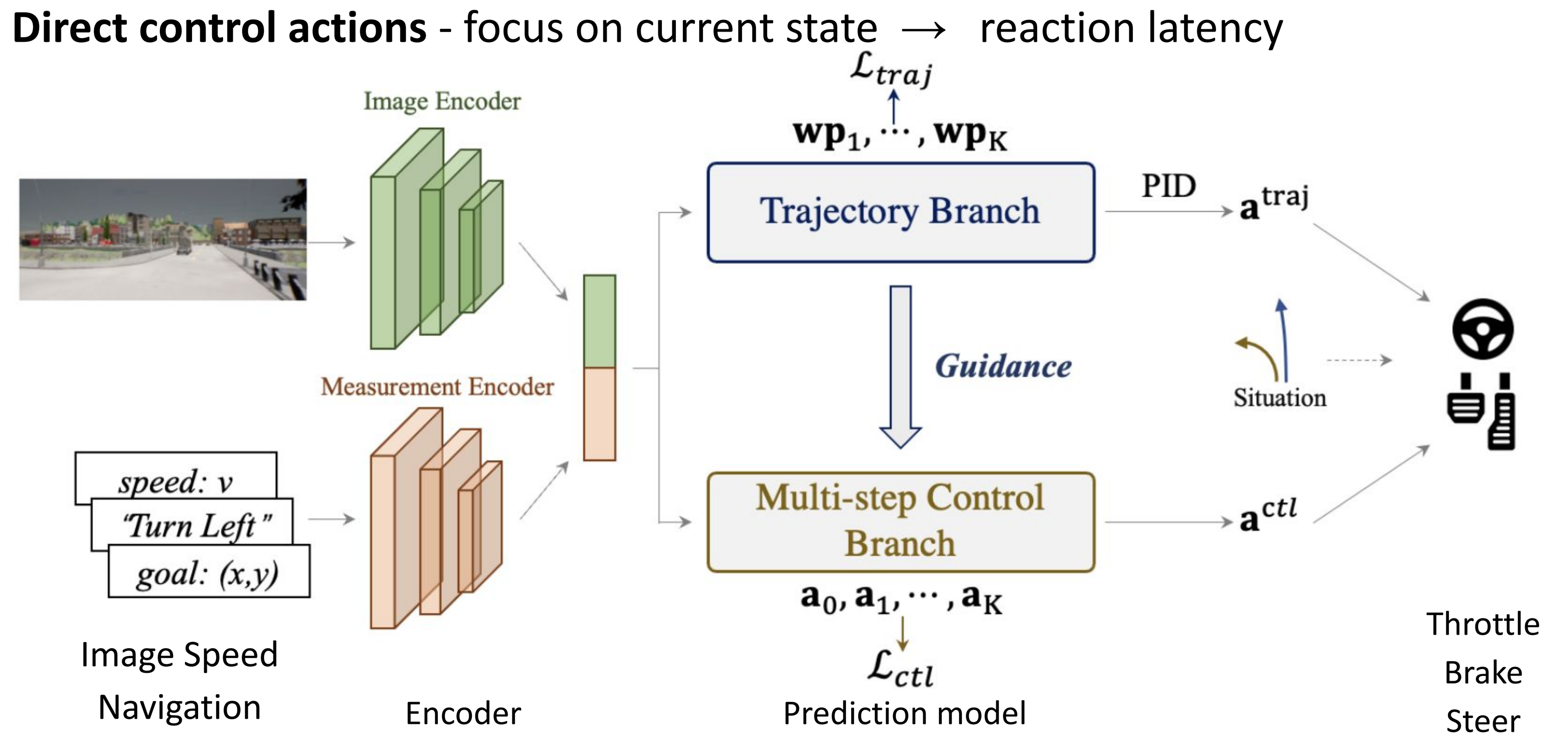

Figure 1. TCP pipeline [2]

- **Trajectory branch:**
  - Expert trajectory as ground truth
  - Supervised training - behavior cloning
  - PID controller for long & lat control
- **Control branch:**
  - Temporal module for multi-step control
  - Trajectory-guided attention map model
  - Imitation learning for control input

- **Output fusion:** Situation based fusion

if *situation* is *trajectory specialized* **then**
  $\mathbf{a} \leftarrow \alpha \times \mathbf{a}^{\mathrm{ctl}} + (1-\alpha) \times \mathbf{a}^{\mathrm{traj}}$
else
  $\mathbf{a} \leftarrow \alpha \times \mathbf{a}^{\mathrm{traj}} + (1-\alpha) \times \mathbf{a}^{\mathrm{ctl}}$
end

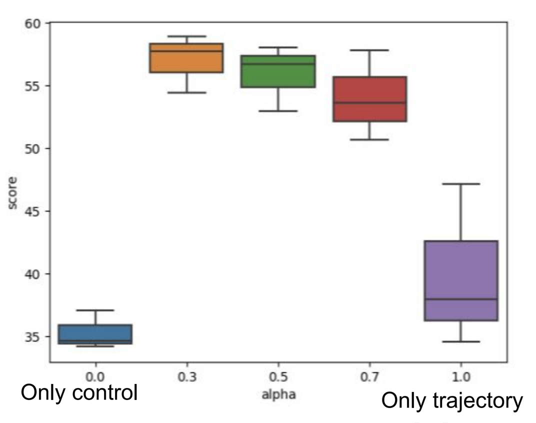

Figure 2. Alpha impact [1]

Highest score in CARLA V1 evaluation (May 2022)

UniAD - Unified Autonomous Driving

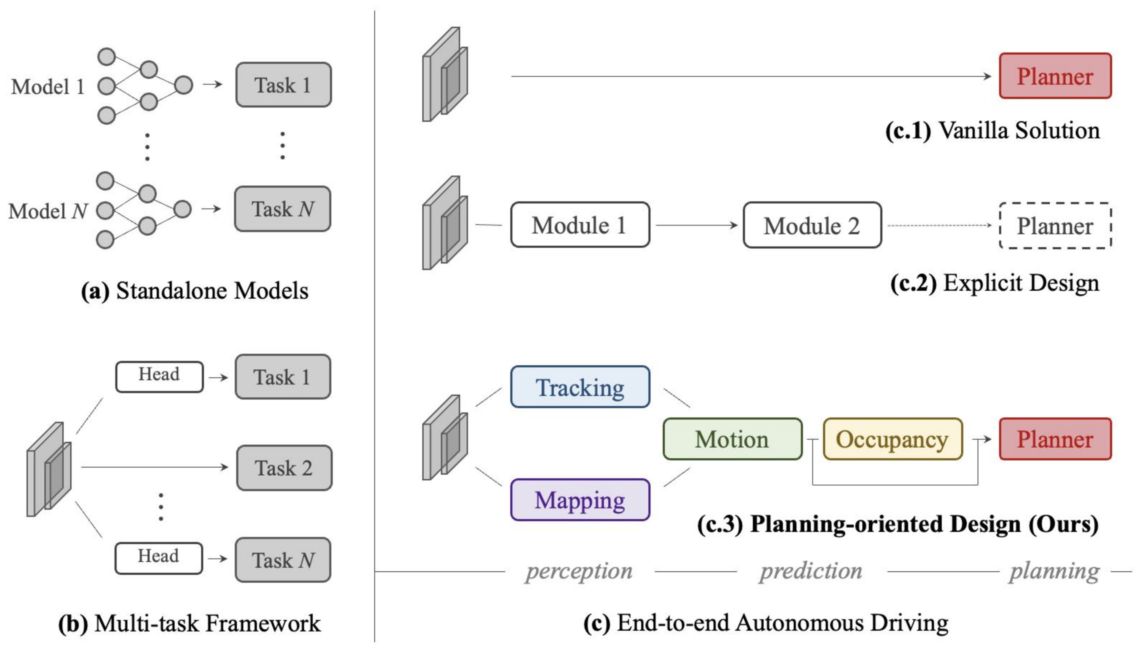

Figure 3. Comparison on various designs of AD framework [2]

- Deploy **standalone models** for individual tasks.
- Design a **multi-task learning (MTL)** paradigm with separate heads.
- Integrate **full-stack** driving tasks into a unified network (**planning-oriented**).

- **Pipeline overview:**

Unified Framework, Bird's-Eye-View (BEV) Features, and Query-Based Design

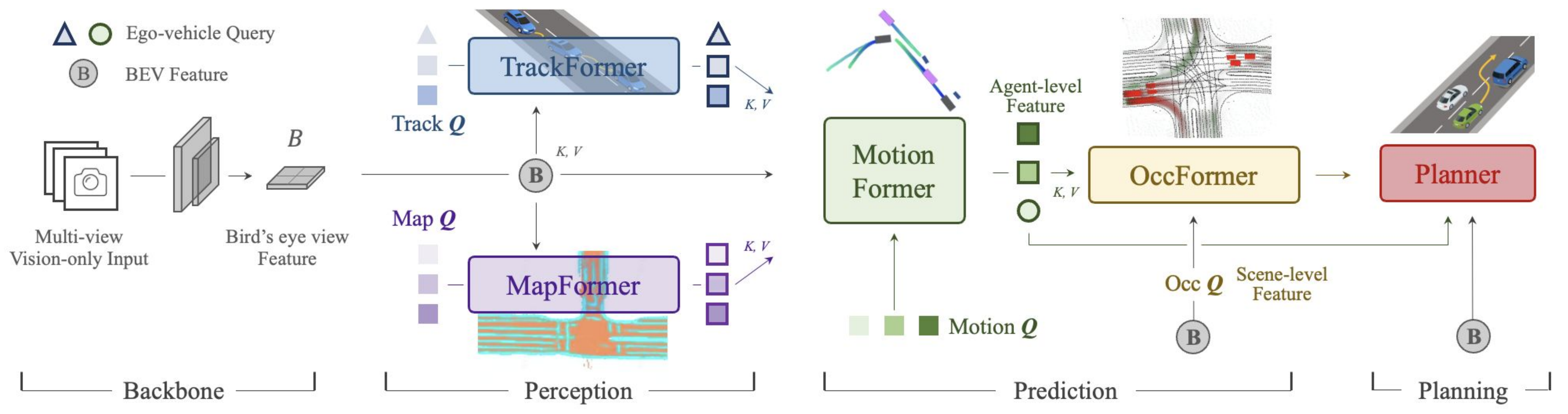

Figure 4. Pipeline of Unified Autonomous Driving (UniAD) [2]

- TrackFormer:
  - Tracks dynamic agents across frames
  - Consistent temporal modeling without post-processing
- MapFormer:
  - Segmentation of road elements
  - Structured scene understanding for planning
- MotionFormer:
  - Multimodal future trajectories for all agents
  - Agent-agent and agent-map interactions
- OccFormer:
  - Instance-level occupancy grids with per-agent identity
  - Scene- and agent-level semantics for collision avoidance

- Planner: generates waypoints and avoids collision using occupancy predictions

$\tau^* = \arg\min_{\tau} f(\tau, \hat{\tau}, \hat{O})$, with the cost function defined as

$$f(\tau, \hat{\tau}, \hat{O}) = \lambda_{\mathrm{coord}} \|\tau, \hat{\tau}\|_2 + \lambda_{\mathrm{obs}} \sum_t \mathcal{D}\left(\tau_t, \hat{O}^t\right)$$

$$\mathcal{D}\left(\tau_t, \hat{O}^t\right) = \sum_{(x,y)\in\mathcal{S}} \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{\|\tau_t - (x,y)\|_2^2}{2\sigma^2}\right)$$

VAD - Vectorized Scene Representation

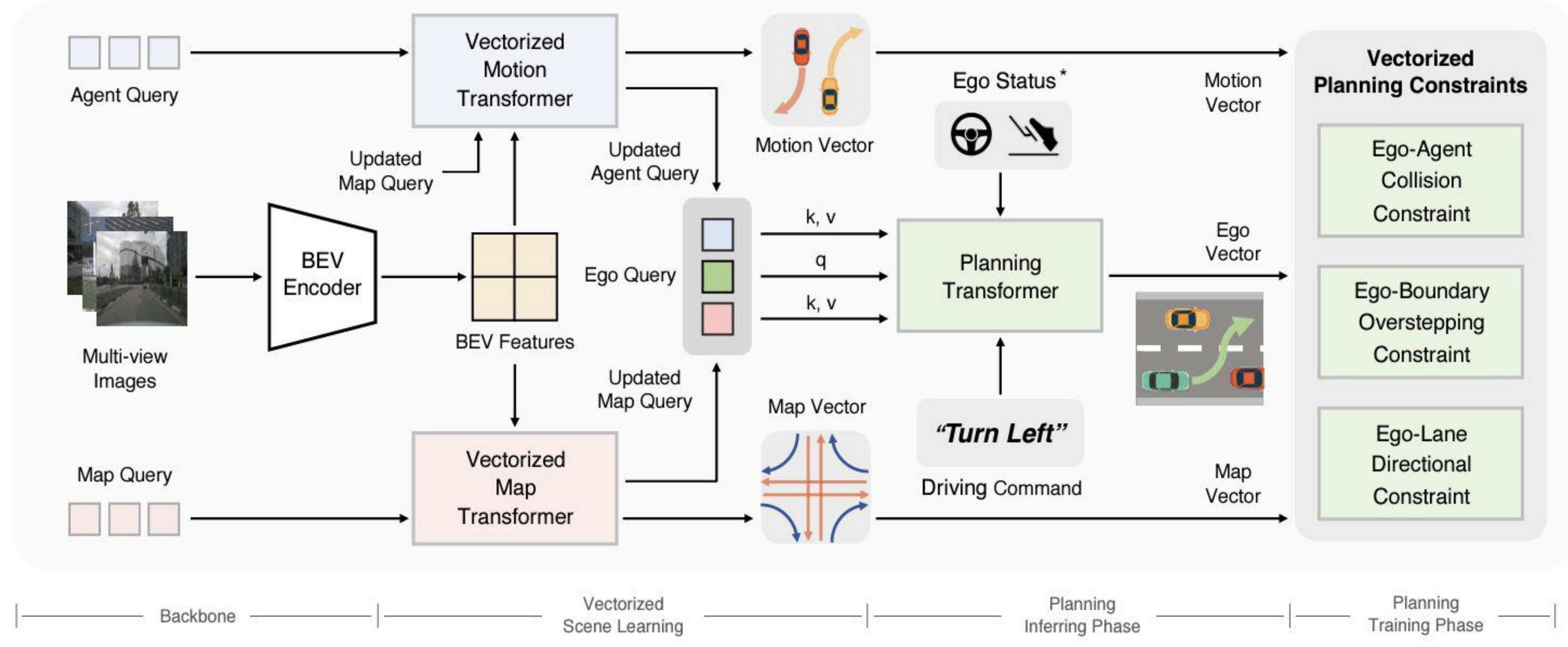

Figure 5. VAD pipeline [3]

- Map:
  - Normalization:
    - Only map, no angle
    - Put every map element under the coordinates system centered on each "target"
  - Filter:
    - Filter the map element that does not have high score
- Query formation:
  - Normally six image input
  - Backbone to decode the input images
  - BEV transformation
  - Get two queries (map/agent)
- Agent:
  - Capture driving intention
  - Interact with dynamic element and static element (self-attention)
  - Cross-attention with map query
  - Predict trajectory information

- Three vectorized planning constraint
  - Ego-agent collision constraint
  - Ego-Boundary Overstepping Constraint
  - Ego-Lane Directional Constraint

Benchmark

Bench2Drive

- The first benchmark for evaluating E2E-AD systems' multiple abilities in a closed-loop manner.

- **Features:**
  - Comprehensive Scenario Coverage
  - Granular Skill Assessment
  - Closed-Loop Evaluation Protocol
  - Diverse Large-Scale Official Training Data

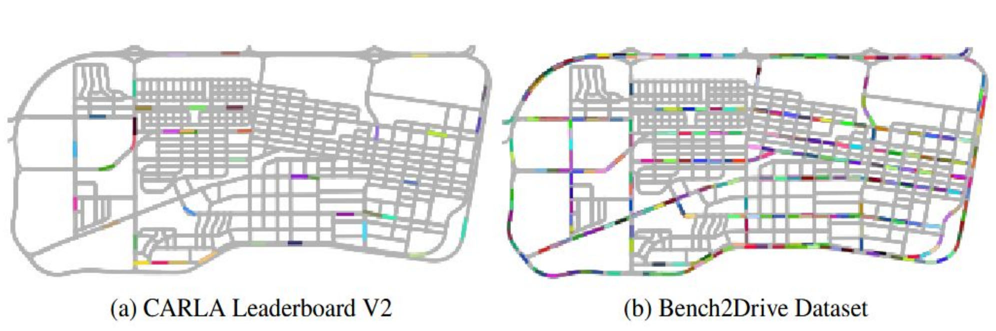

Figure 6. Dataset comparison [4]

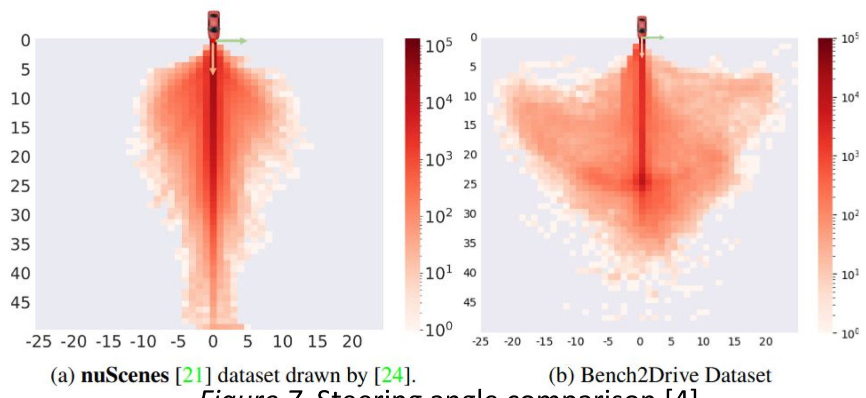

Figure 7. Steering angle comparison [4]

Table 1. Open-loop and Closed-loop Results [4]

Method	Open-loop Metric	Closed-loop Metric			
	Avg. L2 ↓	Driving Score ↑	Success Rate(%) ↑	Efficiency ↑	Comfortness ↑
UniAD-Base [4]	**0.73**	**45.81**	**16.36**	129.21	43.58
VAD [22]	0.91	42.35	15.00	**157.94**	**46.01**
TCP* [5]	1.70	40.70	15.00	54.26	47.80

UniAD gets higher score

Table 2. Multi-Ability Results [4]

Method	Ability (%) ↑				
	Merging	Overtaking	Emergency Brake	Give Way	Traffic Sign
UniAD-Base [4]	14.10	17.78	21.67	10.00	14.21
VAD [22]	8.11	24.44	18.64	20.00	19.15
TCP* [5]	16.18	20.00	20.00	10.00	6.99

UniAD: Emergency Brake
VAD: Overtaking
TCP: Merging

Reference:

[1] Wu, P., Jia, X., Chen, L., Yan, J., Li, H. and Qiao, Y., 2022. Trajectory-guided control prediction for end-to-end autonomous driving: A simple yet strong baseline. *Advances in Neural Information Processing Systems, 35*, pp.6119-6132.

[2] Hu, Y., Yang, J., Chen, L., Li, K., Sima, C., Zhu, X., Chai, S., Du, S., Lin, T., Wang, W., Lu, L., Jia, X., Liu, Q., Dai, J., Qiao, Y., & Li, H. (2023). Planning-oriented Autonomous Driving. arXiv preprint, arXiv:2212.10156.

[3] Jiang, Bo et al. "VAD: Vectorized Scene Representation for Efficient Autonomous Driving." *2023 IEEE/CVF International Conference on Computer Vision (ICCV)* (2023): 8306-8316.

[4] Jia, X., Yang, Z., Li, Q., Zhang, Z. and Yan, J., 2024. Bench2Drive: Towards Multi-Ability Benchmarking of Closed-Loop End-To-End Autonomous Driving. arXiv preprint arXiv:2406.03877.