



TRANSFER YOUR PERSPECTIVE:

Controllable 3D Generation from Any Viewpoint in a Driving Scene

Tai-Yu Pan¹, Sooyoung Jeon¹, Mengdi Fan¹, Jinsu Yoo¹, Zhenyang Feng¹, Mark Campbell², Kilian Q. Weinberger², Bharath Hariharan², Wei-Lun Chao¹

¹The Ohio State University ²Cornell University



TYP (author)

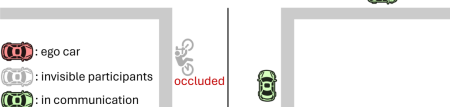


TYP (project)

Highlights

- Introduce a **new research direction**: generating **realistic sensor data** from arbitrary viewpoints to enable **collaborative perception development**
- Propose TYP**: the first solution, successfully transferring single-agent Waymo to collaborative ColWaymo
- Unlock large-scale pre-training** for collaborative perception with **semi-real data**

Future Autonomous Driving



Collaborative Driving:
safer AD system by communicating perception signals or results

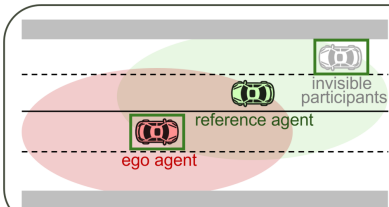
However, **Challenges of Real-World Data Collection for Collaborative Perception Development:**

- $N \times$ more effort than ego-centric setups
- Hard to coordinate multiple vehicles on the same road, same time, within limited range

Thus, existing datasets:

- Primarily simulated
- Few agents (typically only 2)

New Research Direction

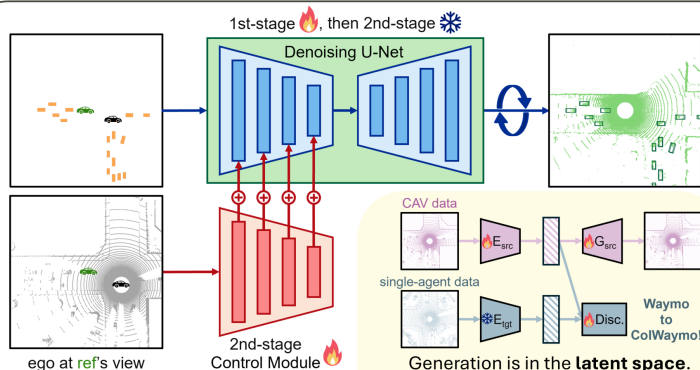


Can we generate 's viewpoint from 's sensory data?

Formally, $P(\mathbf{x}_r | \mathbf{x}_e, \mathbf{y}_r)$,
where \mathbf{y}_r is object locations under ref's view

Problem: We want to convert Waymo into a collaborative dataset—but it only contains ego-view data (\mathbf{x}_e), and lacks reference-view data (\mathbf{x}_r)!

TRANSFER YOUR PERSPECTIVE (TYP)

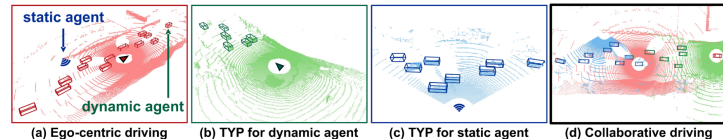


2-stage approach:

- 1st stage: maximize $P(\mathbf{x}|\mathbf{y})$ on Waymo
- 2nd stage: leverage guidance from CAV dataset

Generation is in the **latent space**.
To bridge the **domain gap** between Waymo and the simulated CAV dataset, we incorporate ideas from GANs.

Qualitative Results

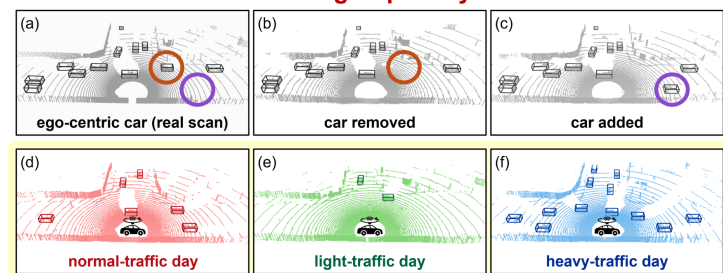


Unlock CAV Pre-Training w/ ColWaymo

Pre-training: scratch / OPV2V / ColWaymo, fine-tuning: V2V4Real, test: V2V4Real

Method	Pre-Train	FT. on 0 Scene				FT. on 5 Scene				FT. on 10 Scene				FT. on 32 Scene			
		s	m	l	all	s	m	l	all	s	m	l	all	s	m	l	all
No Fusion	ego's gt only					0.44	0.16	0.06	0.29	0.66	0.29	0.08	0.42	0.71	0.34	0.08	0.47
	scratch					0.44	0.18	0.16	0.30	0.73	0.28	0.24	0.48	0.79	0.42	0.54	0.60
Early Fusion [9]	OPV2V	0.54	0.20	0.07	0.31	0.32	0.18	0.15	0.25	0.73	0.32	0.31	0.50	0.80	0.47	0.50	0.65
	ColWaymo (ours)	0.50	0.24	0.24	0.35	0.68	0.34	0.32	0.51	0.78	0.39	0.33	0.57	0.83	0.48	0.50	0.65
	scratch	0.44	0.24	0.33	0.34	0.72	0.42	0.51	0.58	0.79	0.53	0.56	0.67				
Late Fusion [80]	OPV2V	0.60	0.27	0.28	0.44	0.55	0.26	0.38	0.42	0.73	0.42	0.51	0.59	0.78	0.55	0.58	0.67
	ColWaymo (ours)	0.40	0.19	0.15	0.25	0.60	0.28	0.44	0.47	0.76	0.43	0.51	0.61	0.82	0.58	0.61	0.71
	scratch	0.40	0.15	0.10	0.28	0.70	0.30	0.17	0.47	0.80	0.45	0.36	0.61				
AttFuse [80]	OPV2V	0.51	0.19	0.05	0.31	0.54	0.22	0.11	0.37	0.77	0.40	0.21	0.54	0.83	0.53	0.40	0.65
	ColWaymo (ours)	0.66	0.35	0.11	0.45	0.65	0.29	0.16	0.46	0.83	0.46	0.33	0.61	0.88	0.58	0.53	0.72
	scratch	0.43	0.15	0.12	0.31	0.70	0.28	0.17	0.43	0.81	0.49	0.30	0.61				
V2X-VIT [79]	OPV2V	0.51	0.24	0.07	0.33	0.48	0.23	0.16	0.35	0.76	0.38	0.22	0.53	0.81	0.49	0.35	0.61
	ColWaymo (ours)	0.60	0.28	0.10	0.34	0.66	0.28	0.22	0.46	0.79	0.48	0.26	0.58	0.84	0.57	0.44	0.67

Editing Capability



Acknowledgement: NSF (IIS-2107077, IIS-2107161), Ohio Supercomputer Center