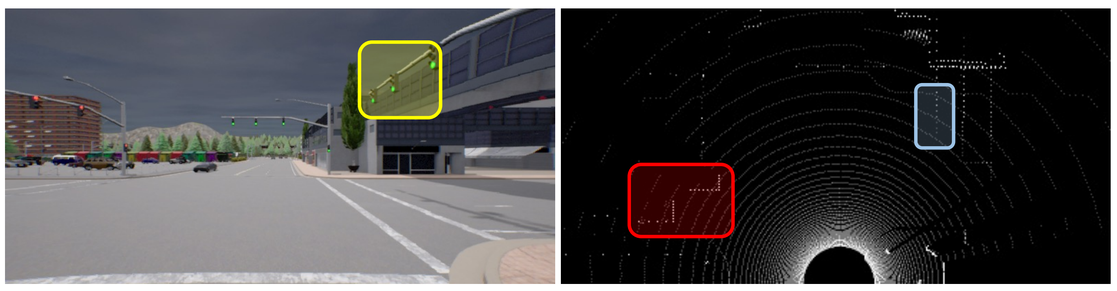
Driving with Attention

Kashyap Chitta[1,2] Aditya Prakash[3] Bernhard Jaeger[1,2]
Zehao Yu[1,2] Katrin Renz[1,2] Andreas Geiger[1,2]

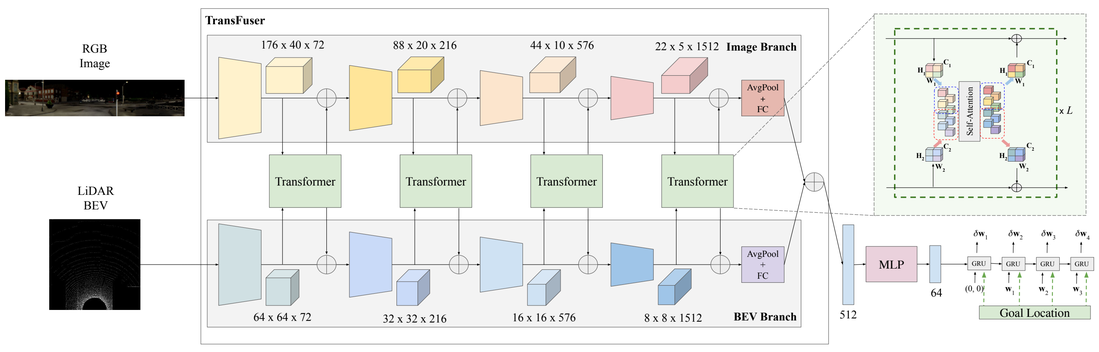
[1]University of Tübingen
[2]Max Planck Institute for Intelligent Systems, Tübingen
[3]University of Illinois Urbana-Champaign

Problem: geometric fusion lacks global context



- Geometric fusion aggregates features from the yellow to the blue region
- However, for safe navigation, it is useful to aggregate features to the red region which contains the vehicles affected by this traffic light

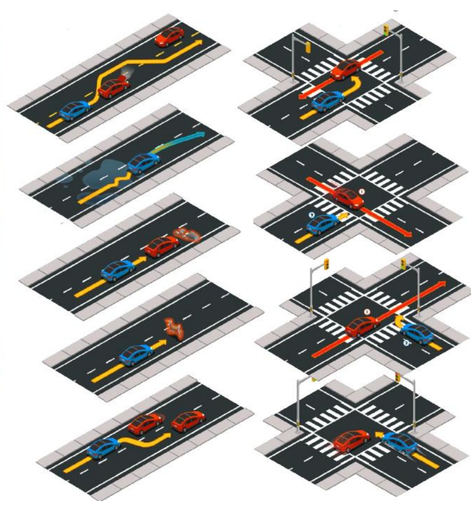
Key idea: attention-based feature fusion



- TransFuser captures the **global context** of the scene **across modalities**
- It uses a simple end-to-end training process based on **imitation learning**

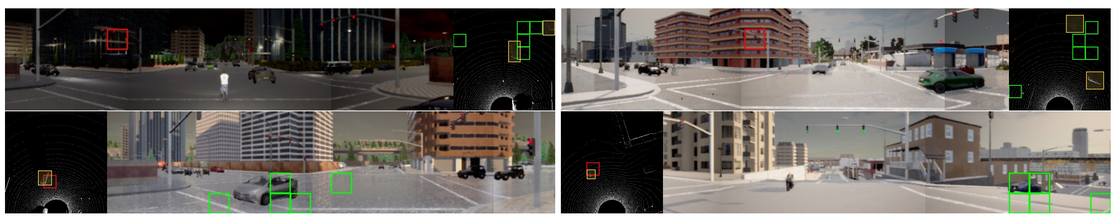
Benchmark: CARLA leaderboard

- 10 routes x 2 weathers x 5 repetitions
- 173 Km of driving experiences



Method	Driving Score ↑	Route Completion ↑	Infraction Score ↑
LAV	61.85	94.46	0.64
TransFuser	**61.18**	**86.69**	**0.71**
Geometric Fusion	41.70	87.85	0.47
GRIAD	36.79	61.85	0.60
WOR	31.37	57.65	0.56

Attention visualizations



EBERHARD KARLS UNIVERSITÄT TÜBINGEN

MAX PLANCK INSTITUTE FOR INTELLIGENT SYSTEMS
imprs-is

Transformers are great for **sensor fusion** in self-driving vehicles!

Take a picture to access the paper