NEAT: Neural Attention Fields for End-to-End Autonomous Driving

MAX PLANCK INSTITUTE FOR INTELLIGENT SYSTEMS

Kashyap Chitta*[1,2] Aditya Prakash*[1] Andreas Geiger[1,2]

[1]Max Planck Institute for Intelligent Systems, Tübingen [2]University of Tübingen

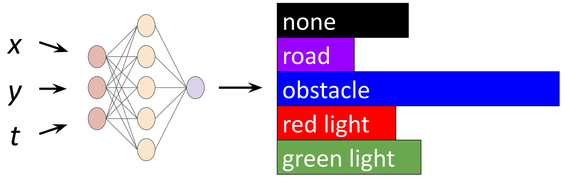
EBERHARD KARLS UNIVERSITÄT TÜBINGEN

Bird's Eye View

- Orthographic projection of 3D world
- No occlusions
- Better correlated with vehicle kinematics

Our Method

Implicit Bird's Eye View Semantics from cameras

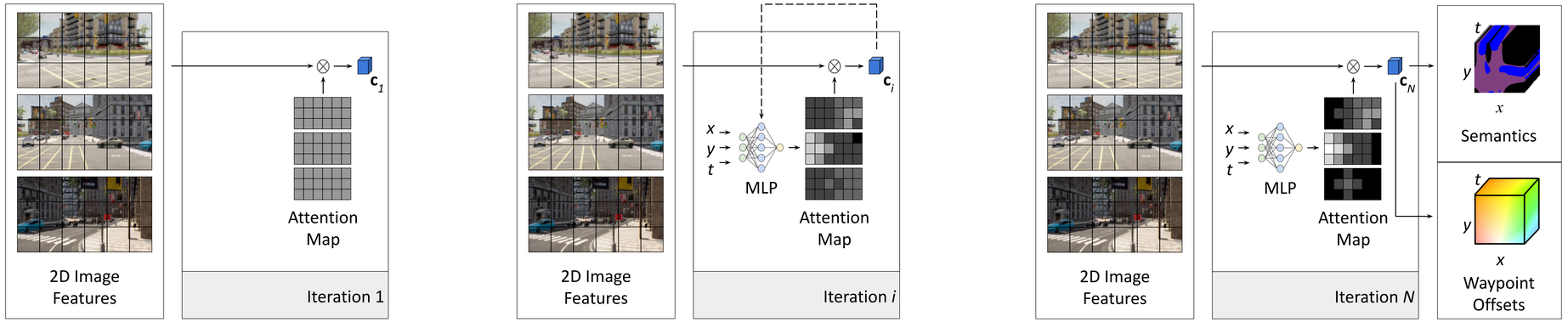


- Compact
- Arbitrary spatial and temporal resolution
- Fixed memory footprint
- Can use sparse supervision

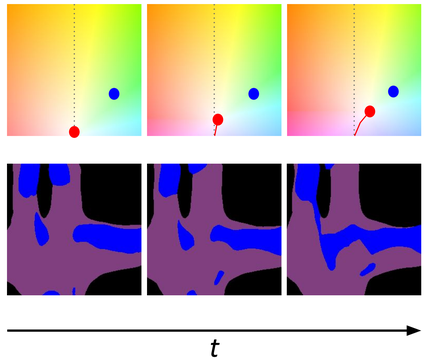
However, associating 2D to BEV semantics requires geometry, scene motion, ego motion, agent intentions and agent interactions

Association by Iterative Refinement of Attention Maps

Uniform prior attention -> location-specific attention maps -> *N* iterations of refinement -> semantic and waypoint offset decoding

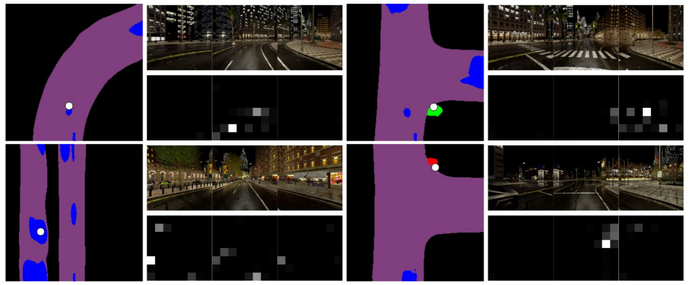


Interpolating Predictions



Top: waypoint offsets (red waypoint, blue target location)
Bottom: predicted semantics

Attention Map Visualizations



Learned attention for highlighted BEV locations (white circle on semantic map). We consistently attend to the object of interest (top left to bottom right: bicyclist, green light, vehicle, red light)

Safe Driving on CARLA Leaderboard

	Team	Submission	Driving score	Route completion	Infraction penalty
		Units	%	%	[0, 1]
⊕	WOR	World on Rails	31.37	57.65	0.56
⊕	MaRLn	MaRLn	24.98	46.97	0.52
⊕	Anonymous	Neural Attention Fields (NEAT)	21.83	41.71	0.65
⊕	Anonymous	CIL-WP	19.38	67.02	0.39
⊕	SDV	TransFuser	16.93	51.82	0.42

(from https://leaderboard.carla.org/ July 2021)

Check out our results video: http://bit.ly/neat_video

kashyap.chitta@tue.mpg.de