

Reading, Writing, and Reviewing

for Robotics and Computer Vision Research

Kashyap Chitta

Autonomous Vision Group

University of Tübingen, Tübingen AI Center

EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN



Agenda

2.1 Reading

2.2 Writing

2.3 Reviewing

2.1

Reading

How do I find relevant literature?

The screenshot shows the Google Scholar search interface. The search bar at the top contains the query "end-to-end autonomous driving". Below the search bar, there are filters: "Articles" selected, "Any time", and "Sort by relevance" (which is underlined). On the left, there are additional filters: "Sort by date", "Any type", "Review articles", and checkboxes for "Include patents" (unchecked), "Include citations" (checked), and "Create alert" (unchecked). The search results list three papers:

- Multimodal end-to-end autonomous driving**
Y. Xiao, F. Codenotti, A. Gurram - IEEE Transactions ..., 2020 - [weexplore.ieee.org](#)
... On the one hand, we find modular pipelines, which divide the driving task into sub-... **end-to-end driving** approaches that try to learn a direct mapping from input raw sensor data to vehicle ...
☆ Save 99 Cite Cited by 135 Related articles All 8 versions
- Multi-modal fusion transformer for end-to-end autonomous driving**
A. Prakash, K. Chittori, A. Geiger - Proceedings of the IEEE ..., 2021 - [openaccess.thecvf.com](#)
... In this work, we propose an architecture for **end-to-end driving** (Fig. 2) with two main components: (1) a MultiModal Fusion Transformer for integrating information from multiple ...
☆ Save 99 Cite Cited by 236 Related articles All 9 versions 30
- End to end learning for self-driving cars**
M. Bojarski, D. Del Testa, D. Dworakowski - arXiv preprint arXiv ..., 2016 - [arxiv.org](#)
... Compared to explicit decomposition of the problem, such as lane marking detection, path planning, and control, our **end-to-end** system optimizes all processing steps simultaneously. ...
☆ Save 99 Cite Cited by 4198 Related articles All 15 versions 30

Each paper entry includes a link to the full text: "[PDF] [arxiv.org](#)" or "[PDF] [thecvf.com](#)". The last two entries are highlighted with a red border.

<https://scholar.google.com/>

- ▶ No starting point (=paper) → use **keyword** based search
- ▶ Luckily, most papers are **open access** in vision/robotics

Paper as starting point

Multi-modal fusion transformer for **end-to-end autonomous driving**

[A Prakash, K Chitta, A Geiger - Proceedings of the IEEE ..., 2021 - openaccess.thecvf.com](#)

... In this work, we propose an architecture for **end-to-end driving** (Fig. 2) with two main components: (1) a MultiModal Fusion Transformer for integrating information from multiple ...

☆ Save ⟲ Cite **Cited by 236** Related articles All 9 versions ⟳

- ▶ **Forward** and **backward** search
- ▶ For **finding newer papers** → "Cited by" on Google Scholar
- ▶ Can sort the results by relevance or date
- ▶ Can also restrict the time range for which papers are displayed

Backward search

2. Related Work

Multi-Modal Autonomous Driving: Recent multi-modal methods for end-to-end driving [58, 65, 51, 3] have shown that complementing RGB images with depth and semantics has the potential to improve driving performance. Xiao et al. [58] explore RGBD input from the perspective of early, mid and late fusion of camera and depth modalities and ob-

we propose an attention-based Multi-Modal Fusion Transformer that incorporates global contextual reasoning and achieves superior driving performance.

Attention for Autonomous Driving: Attention has been explored in the context of driving for lane changing [13], object detection [11, 32] and motion forecasting [32, 50, 49, 28, 15, 30, 29, 56]. Chen et al. [11] employ a recurrent attention mechanism over a learned semantic map for

References

- [1] Waymo open dataset: An autonomous driving dataset. <https://www.waymo.com/open>, 2019.
- [2] Mayank Bansal, Alex Krizhevsky, and Abhijit S. Ogale. Chauffeurnet: Learning to drive by imitating the best and synthesizing the worst. In *Proc. Robotics: Science and Systems (RSS)*, 2019.
- [3] Aseem Behl, Kashyap Chitta, Aditya Prakash, Eshed Ohn-Bar, and Andreas Geiger. Label efficient visual abstractions for autonomous driving. In *Proc. IEEE International Conf. on Intelligent Robots and Systems (IROS)*, 2020.
- [15] Chiho Choi and Behzad Dariush. Looking to relations for future trajectory forecast. In *Proc. of the IEEE International Conf. on Computer Vision (ICCV)*, 2019.
- [16] Felipe Codevilla, Eder Santana, Antonio M. López, and Adrien Gaidon. Exploring the limitations of behavior cloning for autonomous driving. In *Proc. of the IEEE International Conf. on Computer Vision (ICCV)*, 2019.
- [17] Marius Cordts, Mohamed Omran, Sebastian Ramos, Timo Rehfeld, Markus Enzweiler, Rodrigo Benenson, Uwe Franke, Stefan Roth, and Bernt Schiele. The cityscapes dataset for semantic urban scene understanding. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2016.

- For **finding older papers** → "Related Work" section
- Search title via Google Scholar

Track the people!

Kashyap Chitta

I am a PhD student at the University of Tübingen, Germany, where I am part of the [Autonomous Vision Group](#) led by Prof. Andreas Geiger. *I plan to finish my PhD by the end of 2023 and I am looking for postdoc positions!*

News: I was recently selected as an [RSS pioneer](#) for 2023, and nominated as an [outstanding reviewer](#) for CVPR 2023. Our team also won the two most recent closed-loop driving challenges: the 2022 CARLA autonomous driving challenge (map track) and 2023 nPlan planning challenge!

Research: I am excited about data-driven solutions to complex decision-making tasks. Currently, my research focuses on self-driving vehicles. Specifically, I am interested in how autonomous agents can use attention-based deep neural networks to create abstract representations suitable for safe navigation. Further, I am a big fan of simulation, and am interested in building data-driven simulators tailored towards improving the robustness and generalization of learned policies. Representative papers are highlighted below.

Bio: Kashyap did a bachelor's degree in electronics at the RV College of Engineering, India. He then moved to the US in 2017 to obtain his Master's degree in computer vision from Carnegie Mellon University, where he was advised by Prof. Martial Hebert. During this time, he was also an intern at NVIDIAGA, working with Dr. Juan M. Alvarado. He is currently a PhD student in the Autonomous Vision Group at the University of Tübingen, Germany, supervised by Prof. Andreas Geiger.

[CV](#) [Mall](#) [Scholar](#) [Twitter](#) [LinkedIn](#) [Facebook](#) [Mastodon](#) [GitHub](#) [YouTube](#)



Publications



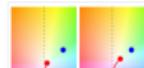
Plant: Explainable Planning Transformers via Object-Level Representations
Kaihui Han, Kashyap Chitta, Oskar-Martin Meier, Soparna Koenke, Zeynep Akata, Andreas Geiger
Conference on Robot Learning (CoRL), 2022
[Abs](#) / [Paper](#) / [Supplementary](#) / [Video](#) / [Code](#) / [Bibtex](#)



KND: Generating Safety-Critical Driving Scenarios for Robust Imitation via Kinematics Gradients (Oral)
Niklas Hausemann, Kathrin Renz, Kashyap Chitta, Aparna Bhattacharyya, Andreas Geiger
European Conference on Computer Vision (ECCV), 2022
[Abs](#) / [Paper](#) / [Supplementary](#) / [Video](#) / [Poster](#) / [Code](#) / [Bibtex](#)



TransFuser: Imitation with Transformer-Based Sensor Fusion for Autonomous Driving
Kashyap Chitta, Aditya Prakash, Bernhard Jaeger, Zehan Yu, Kathrin Renz, Andreas Geiger
Transactions on Pattern Analysis and Machine Intelligence (T-PAMI), 2022
[Abs](#) / [Paper](#) / [Supplementary](#) / [Video](#) / [Poster](#) / [Code](#) / [Bibtex](#)



NEAT: Neural Attention Fields for End-to-End Autonomous Driving
Kashyap Chitta, Aditya Prakash, Andreas Geiger
International Conference on Computer Vision (ICCV), 2021
[Abs](#) / [Paper](#) / [Supplementary](#) / [Video](#) / [Poster](#) / [Code](#) / [Bibtex](#)



Benchmarking Unsupervised Object Representations for Video Sequences
Manisha Iwasiuk, Kashyap Chitta, Yash Sharma, Michael Brandel, Matthias Bethge, Andreas Geiger, Alexander Ecker
Journal of Machine Learning Research (JMLR), 2021
[Abs](#) / [Paper](#) / [Video](#) / [Code](#) / [Bibtex](#)



Multi-Modal Fusion Transformer for End-to-End Autonomous Driving
Aditya Prakash, Kashyap Chitta, Andreas Geiger
Conference on Computer Vision and Pattern Recognition (CVPR), 2021
[Abs](#) / [Paper](#) / [Supplementary](#) / [Video](#) / [Poster](#) / [Code](#) / [Bibtex](#)



Label Efficient Visual Abstractions for Autonomous Driving
Aseem Behl, Kashyap Chitta, Aditya Prakash, Eshed Ohn-Bar, Andreas Geiger
International Conference on Intelligent Robots and Systems (IROS), 2020
[Abs](#) / [Paper](#) / [Video](#) / [Bibtex](#)



Scalable Active Learning for Object Detection
Elmar Haussmann, Michele Fenzi, Kashyap Chitta, Jan Ivanekcy, Harrison Xu, Dorna Roy, Akhila Mittel, Nicolas Kouroufatzikis, Clement Farabet, Jose Alvarez
Intelligent Vehicles Symposium (IV), 2020
[Abs](#) / [Paper](#) / [Bibtex](#)



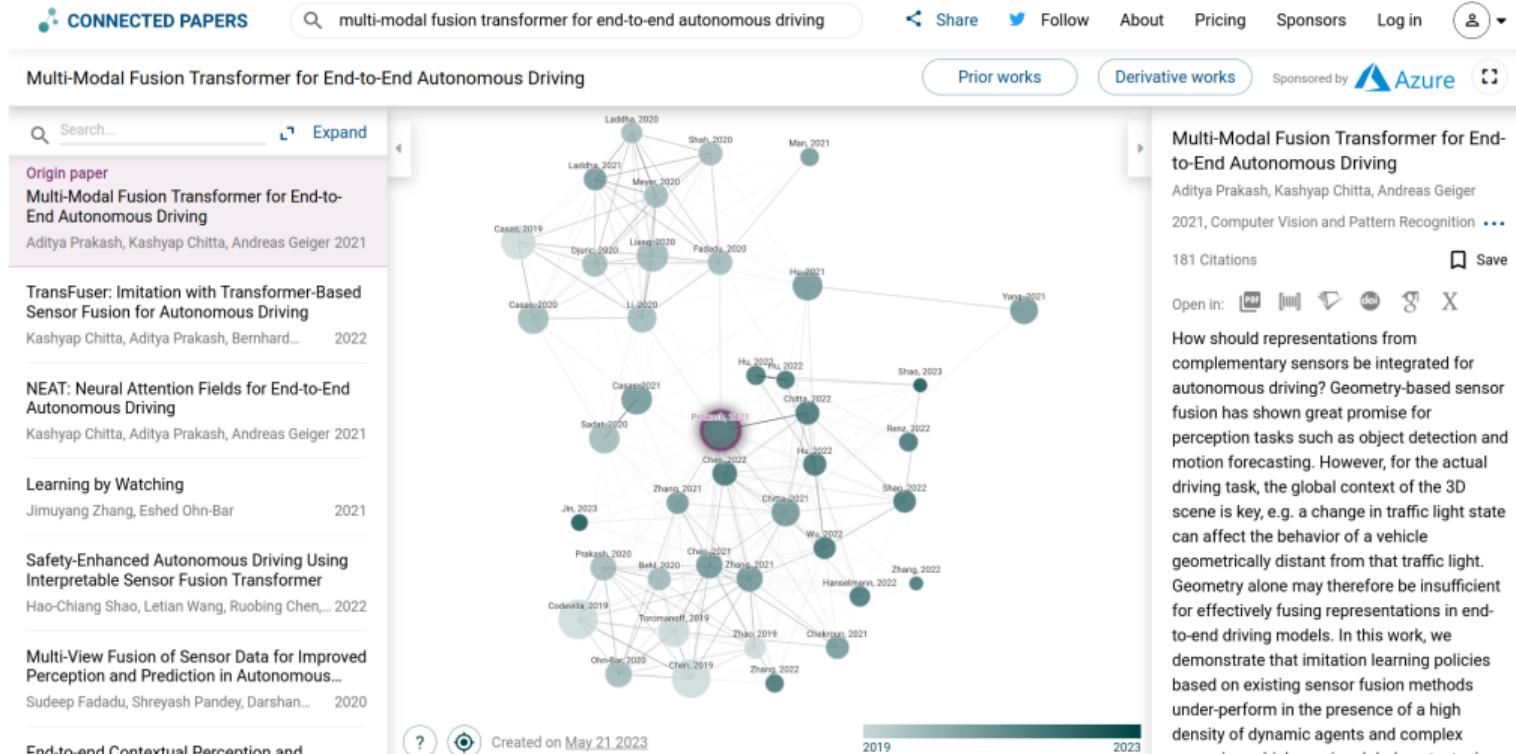
Learning Situational Driving
Eshed Ohn-Bar, Aditya Prakash, Aseem Behl, Kashyap Chitta, Andreas Geiger
Conference on Computer Vision and Pattern Recognition (CVPR), 2020
[Abs](#) / [Paper](#) / [Supplementary](#) / [Video](#) / [Bibtex](#)



Exploring Data Aggregation in Policy Learning for Vision-Based Urban Autonomous Driving
Aditya Prakash, Aseem Behl, Eshed Ohn-Bar, Kashyap Chitta, Andreas Geiger
Conference on Computer Vision and Pattern Recognition (CVPR), 2020
[Abs](#) / [Paper](#) / [Supplementary](#) / [Video](#) / [Code](#) / [Bibtex](#)

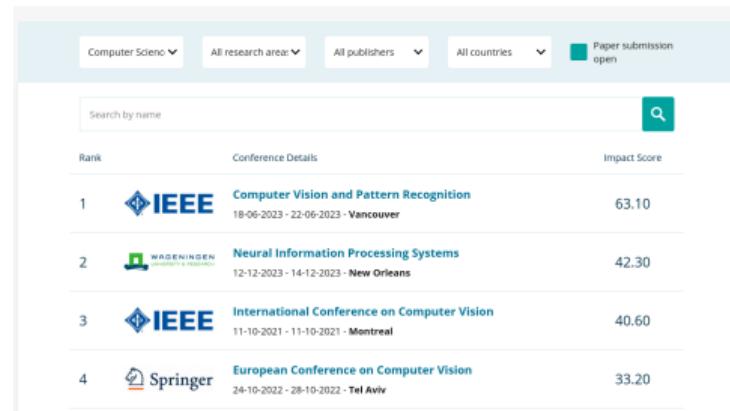
- Find **authors' websites**: often the same author has written other related work

Specialized tools: paper graphs



Is a paper worth reading?

- ▶ You must **invest time**, it is hard work
- ▶ But you can't read all papers in depth
- ▶ Read **abstract** and look at **teaser figure**
- ▶ Look at **impact** of paper (citations) and conference/journal
- ▶ Top conferences are selective and have acceptance rate of 25% or lower
- ▶ **Rankings:** <https://research.com/>



The screenshot shows a search interface for conference rankings. The search bar contains 'Computer Science'. Below the search bar are filters: 'All research area', 'All publishers', 'All countries', and a checked filter 'Paper submission open'. A search button with a magnifying glass icon is also present. The main table lists four conferences with their details and impact scores:

Rank	Conference Details	Impact Score
1	 IEEE Computer Vision and Pattern Recognition 18-06-2023 - 22-06-2023 - Vancouver	63.10
2	 WAGENINGEN UNIVERSITY & RESEARCH Neural Information Processing Systems 12-12-2023 - 14-12-2023 - New Orleans	42.30
3	 IEEE International Conference on Computer Vision 11-10-2021 - 11-10-2021 - Montreal	40.60
4	 Springer European Conference on Computer Vision 24-10-2022 - 28-10-2022 - Tel Aviv	33.20

Know the state-of-the-art

- Good methods should **perform** well
- **Benchmarks** have established as an important tool to measure progress
- Benchmarks often link papers and code

CARLA Leaderboard 1.0 – SENSORS Track

Copy CSV

Team	Submission	Driving score	Route completion	Infraction penalty
Units	%	%	[0, 1]	
Interfuser	ReasonNet	79.95	89.89	0.89
Interfuser	InterFuser	76.18	88.23	0.84
PPX	TCP	75.14	85.63	0.87
WOR	Learning from All Vehicles (LAV)	61.85	94.46	0.64
DP	TransFuser	61.18	86.69	0.71
Attention Fields	Latent TransFuser	45.20	66.31	0.72
Raphael	General Reinforced Imitation for Autonomous Driving (GRIAD)	36.79	61.85	0.60
DP	TransFuser+	34.58	69.84	0.56
WOR	World on Rails	31.37	57.65	0.56
MaRLn	MaRLn	24.98	46.97	0.52

Showing 1 to 10 of 23 entries

<https://leaderboard.carla.org>

How to read a paper?

Abstract

How should representations from complementary sensors be integrated for autonomous driving? Geometry-based sensor fusion has shown great promise for perception tasks such as object detection and motion forecasting. However, for the actual driving task, the global context of the 3D scene is key, e.g., a change in traffic light state can affect the behavior of a vehicle geometrically distant from that traffic light. Geometry alone may therefore be insufficient for effectively fusing representations in end-to-end driving models. In this work, we demonstrate that imitation learning policies based on existing sensor fusion methods under-perform in the presence of a high density of dynamic agents and complex scenarios, which require global contextual reasoning, such as handling traffic oncoming from multiple directions at uncontrolled intersections. Therefore, we propose TransFuser, a novel Multi-Modal Fusion Transformer, to integrate image and LiDAR representations using attention. We experimentally validate the efficacy of our approach in urban settings involving complex scenarios using the CARLA urban driving simulator. Our approach achieves state-of-the-art driving performance while reducing collisions by 76% compared to geometry-based fusion.

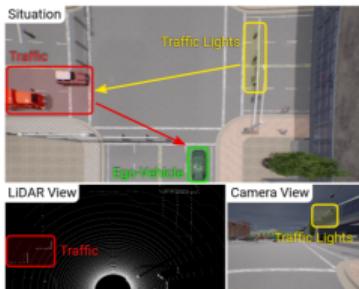
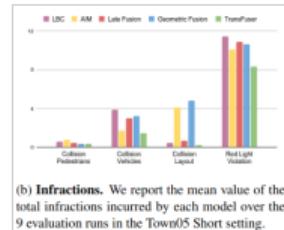


Figure 1: **Illustration.** Consider an intersection with oncoming traffic from the left. To safely navigate the intersection, the ego-vehicle (green) must capture the global context of the scene involving the interaction between the traffic light (yellow) and the vehicles (red). However, the traffic light state is not visible in the LiDAR point cloud and the vehicles are not visible in the camera view. Our TransFuser model integrates both modalities via global attention mechanisms to capture the 3D context and navigate safely.

Contributions: (1) We demonstrate that imitation learning policies based on existing sensor fusion approaches are unable to handle adversarial scenarios in urban driving, e.g., unprotected turnings at intersections or pedestrians emerging from occluded regions. (2) We propose a novel Multi-Modal Fusion Transformer (TransFuser) to incorporate the global context of the 3D scene into the feature extraction layers of different modalities. (3) We experimentally validate our approach in complex urban settings involving adversarial scenarios in CARLA and achieve state-of-the-art performance. Our code and trained models are available at <https://github.com/autonomousvision/transfuser>.

2. Related Work



- Unless 100% sure the paper is relevant, **don't** read it linearly from start to end
- Instead, take a **quick look** at abstract, teaser, contributions, results (~ 10 min)
- Take notes, summarize, and decide if you want to **read it in depth** (2+h)

Keep notes

- ▶ Pdf annotation → Okular, Acrobat, Mendeley ...
- ▶ **Highlight** important lines
 - ▶ Important (blue)
(design choices, models, results)
 - ▶ Interesting (green)
 - ▶ Confusing (yellow)
 - ▶ Suspicious (red)

DiffStack: A Differentiable *and* Modular Control Stack for Autonomous Vehicles

Peter Karkus¹, Boris Ivanovic¹, Shie Mannor^{1,2}, Marco Pavone^{1,3}

¹NVIDIA Research, ²Techion, ³Stanford University
{pkarkus,bivanovic,smannor,mpavone}@nvidia.com

Abstract: Autonomous vehicle (AV) stacks are typically built in a modular fashion, with explicit components performing detection, tracking, prediction, planning, control, etc. While modularity improves reusability, interpretability, and generalizability, it also suffers from compounding errors, information bottlenecks, and integration challenges. To overcome these challenges, a prominent approach is to convert the AV stack into an end-to-end neural network and train it with data. While such approaches have achieved impressive results, they typically lack interpretability and reusability, and they eschew principled analytical components, such as planning and control, in favor of deep neural networks. To enable the joint optimization of AV stacks while retaining modularity, we present DiffStack, a differentiable *and* modular stack for prediction, planning, and control. Crucially, our model-based planning and control algorithms leverage recent advancements in differentiable optimization to produce gradients, enabling optimization of upstream components, such as prediction, via backpropagation through planning and control. Our results on the nuScenes dataset indicate that end-to-end training with DiffStack yields substantial improvements in open-loop and closed-loop planning metrics by, e.g., learning to make fewer prediction errors that would affect planning. Beyond these immediate benefits, DiffStack opens up new opportunities for fully data-driven yet modular and interpretable AV architectures. Project website: <https://sites.google.com/view/diffstack>

Keywords: Differentiable Algorithms, Autonomous Driving, Planning, Control.

1 Introduction

Intelligent robotic systems, such as autonomous vehicles (AVs), are typically architected in a modular fashion and comprised of modules performing detection, tracking, prediction, planning, and control, among others [1, 2, 3, 4, 5, 6, 7, 8]. Modular architectures are generally desirable because of

www.mendeley.com

Look up unknown concepts



$\frac{\delta \mathcal{L}_{\text{plan}}}{\delta w} = \frac{\delta \mathcal{L}_{\text{CE}}}{\delta p_n} \frac{\delta p_n}{\delta C} \frac{\delta C}{\delta w}$; and similarly, $\frac{\delta \mathcal{L}_{\text{plan}}}{\delta \theta} = \frac{\delta \mathcal{L}_{\text{CE}}}{\delta p_n} \frac{\delta p_n}{\delta C} \frac{\delta C}{\delta C_{\text{coll}}} \frac{\delta C_{\text{coll}}}{\hat{s}_a} \frac{\hat{s}_a}{\theta}$, where all terms exist.

Control. The control module performs MPC over a finite horizon using an iterative box-constrained linear quadratic regulator (LQR) algorithm [43]. Formally, we aim to solve

$$s_{\text{ctr}}, u_{\text{ctr}} = \arg \min_{s, u} C(s, u; \hat{s}_a \in A, g, m; w) \quad \text{s.t. } s^{(0)} = s^{\text{init}}, s^{(t+1)} = f_d(s^{(t)}, u^{(t)}), \underline{u} \leq u \leq \bar{u}, \quad (4)$$

where C denotes the cost function, f_d the dynamics, s^{init} the current ego state, and \underline{u}, \bar{u} the control limits. We use the cost defined in (3) for C and the dynamically-extended unicycle [44] for f_d . We initialize the trajectory with u_{plan} from the planner. The algorithm then iteratively forms and solves a quadratic LQR approximation of (4) around the current solution $s^{(i)}, u^{(i)}$ for iteration i , using first- and second-order Taylor approximations of f_d and C , respectively. The trajectory is updated to be close to the LQR optimal control while also decreasing the original non-quadratic cost. We stop iterations upon convergence or a fixed limit.

Linear–quadratic regulator

8 languages ▾

Article Talk

Read Edit View history Tools ▾

From Wikipedia, the free encyclopedia

The theory of [optimal control](#) is concerned with operating a [dynamic system](#) at minimum cost. The case where the system dynamics are described by a set of [linear differential equations](#) and the cost is described by a [quadratic function](#) is called the LQ problem. One of the main results in the theory is that the solution is provided by the **linear-quadratic regulator (LQR)**, a feedback controller whose equations are given below.

LQR controllers possess inherent robustness with guaranteed [gain](#) and [phase margin](#),^[1] and they also are part of the solution to the LQG (linear-quadratic-Gaussian) problem. Like the LQR problem itself, the LQG problem is one of the most fundamental problems in control theory.

Read prior work when necessary



3.1 DiffStack modules

Prediction. We employ Trajectron++ [42], a state-of-the art CVAE that takes H seconds of state history for all agents as input, and outputs multimodal trajectory predictions for one agent $a \in A$,

$$\hat{s}_a^{(1:T)}(\theta) = \{\hat{s}_{a,k}^{(1:T)}(\theta)\}_{k \in K} = \text{CVAE}\left(s_{a' \in A}^{(-H:0)}; \theta\right), \quad (1)$$

where $k \in K$ is the mode of the output distribution. We will use $\hat{s}_a = \hat{s}_a^{(1:T)}(\theta)$ for brevity. The encoder of the CVAE processes agent state histories with recurrent LSTM networks and models inter-agent interactions using graph-based attention. The decoder is a GRU that outputs a Gaussian Mixture Model (GMM) for each future timestep. The GMM modes correspond to the CVAE's $K = 25$ discrete latent states. To ensure predictions are dynamically-feasible, GMMs are defined over controls and then integrated through a known (differentiable) dynamics function to produce a trajectory. We use the default model configuration without map and ego conditioning. We augment the input states with an ego-indicator variable to allow for ego-agent relation reasoning. The raw prediction training objective is the InfoVAE loss, $\mathcal{L}_{\text{pred}} = \mathcal{L}_{\text{InfoVAE}}(\hat{s}_a, s_a^{\text{gt}})$, the same as for the original Trajectron++.

Trajectron++: Dynamically-Feasible Trajectory Forecasting With Heterogeneous Data

Tim Salzmann^{*†1}, Boris Ivanovic^{*1}, Punarjay Chakravarty², and Marco Pavone¹

¹ Autonomous Systems Lab, Stanford University

{timsal, borisi, pavone}@stanford.edu

² Ford Greenfield Labs

pchakra5@ford.com

Use project pages

- ▶ Often contain:
 - ▶ Talks and **slides**
 - ▶ Narrated **videos**
 - ▶ Supplementary materials
 - ▶ Source code (e.g., github)
 - ▶ Additional resources (e.g., blog)
- ▶ Use these resources to quickly get a high-level understanding of a paper
- ▶ For more tips on reading, see **Jia-Bin Huang's thread** linked in the footnote

Multi-Modal Fusion Transformer for End-to-End Autonomous Driving

CVPR 2021

Aditya Prakash^{*} Kashyap Chitta^{*} Andreas Geiger

Max Planck Institute for Intelligent Systems University of Tübingen



[Paper] [Supplementary] [Code] [Date] [Video] [Poster] [Blog]

Abstract

How should representations from complementary sensors be integrated for autonomous driving? While depth and radar are well suited for object detection and motion forecasting, however, for the actual driving task, the global context of the scene is equally important. In this work, we show that the sensor fusion of a vehicle geometrically deviates from that traffic lights. Geometry alone may therefore be insufficient for effectively fusing representations, as end-to-end learning has shown. We propose a multi-modal fusion transformer that fuses the outputs of existing sensor fusion methods under-perform in the presence of a high density of dynamic objects. Our approach is based on a multi-head attention mechanism. As an example, we demonstrate traffic sensing from multiple directions at uncontrolled intersections. Furthermore, we propose a transformer-based multi-modal fusion framework to incorporate these learned representations into a neural network. We experimentally validate the efficiency of our approach in urban settings involving complex scenarios using the CMU urban driving dataset. Our approach achieves a 10% improvement in the driving performance while reducing collisions by 30% compared to geometry-based fusion.

Video



Generalization to New Town



Which tool can I use to write?

- Our community → \LaTeX , standard tool for **academic typesetting**
 - Professional typesetting of text, equations, figures and tables

Overleaf

The screenshot shows the Overleaf LaTeX editor interface. On the left, the file structure is visible with files like `imgtx`, `listab`, `bibliography_strings.bib`, `bibliography.bib`, `cvpr.sty`, `leso-pic.sty`, `leeebst`, `shortcuts.tex`, and the main file `top.tex` which is currently selected. The `top.tex` code includes LaTeX commands for packages, document structure, and bibliography. A sidebar on the left shows a 'File outline' with sections: Introduction, Methods, Evaluation, and Discussion. The main workspace displays the LaTeX code and its corresponding PDF preview. The PDF preview shows a title page with 'Your Title' and '3DV Seminar Report'. The right side of the interface contains a vertical navigation bar with file names and line numbers from 000 to 060. At the top, there are buttons for 'Review', 'Share', 'Submit', 'History', 'Chat', and 'View warning'.

www.overleaf.com

- Online, no installation, good for beginners

How can I learn L^AT_EX?

The screenshot shows the Overleaf Documentation homepage. At the top, there's a navigation bar with links for "Features & Benefits", "Templates", "Help", "Projects", and "Account". Below the navigation is a search bar with the placeholder "Search help library...". The main content area has a title "Documentation" and a sub-section "New to LaTeX?". It includes a brief welcome message and a list of introductory topics. To the left, there's a sidebar with a tree-like navigation menu. The menu categories include "Documentation Home", "Overleaf guides", and "LaTeX Basics". Under "Overleaf guides", there are links for creating documents, uploading projects, copying projects, and various project management tasks. Under "LaTeX Basics", there are links for creating first LaTeX documents, choosing compilers, and basic typesetting features like bold, italic, and underline.

www.overleaf.com/learn

- **Overleaf documentation** provides great resources

Start early and iterate!

- ▶ Writing needs **time**
- ▶ **Ideas** form while writing
- ▶ **Problems** surface while writing
- ▶ It is important to start writing **early** on and iterate
- ▶ Bullets → long text → concise text
- ▶ Get feedback!

Come up with a good structure

- ▶ **Abstract** – Task, challenge, idea, result (200-400 words)
- ▶ **Introduction**
 - ▶ Definition – What is the problem? Where does it occur?
 - ▶ Motivation – Why should we care? What applications?
 - ▶ Contributions – What is now possible as a result of your work? Why was this not possible before?
- ▶ **Related Work** – What has been done? How are you different?
- ▶ **Method** – How does it work? Why design the system this way?
- ▶ **Results** – What has been achieved? What works and what doesn't? Why?
- ▶ **Conclusion** – What should we have learned? Limitations? Future work?

Equations should remove ambiguity

- ▶ Formalize using math **when appropriate**
- ▶ Introduce every mathematical symbol that you are using
- ▶ Provide **intuitions** wherever possible
- ▶ Be as concise but **precise**
- ▶ Redundancy is fine for key concepts! (e.g. equation + figure + text)

Figures help understanding

- ▶ Place figures outside running text, usually at **top of page**
- ▶ Adjust figure font size to font size of main text
- ▶ **Caption** should describe figure concisely to be understood stand-alone
- ▶ When using a figure or table from another source, **cite** the source in the caption
- ▶ Make sure all figures and tables are **referenced** from the main text
- ▶ You can reference the same figure or table multiple times

Minimize white space

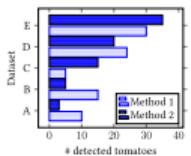


Figure 1: The number of tomatoes detected by either of the two methods across the five main datasets.

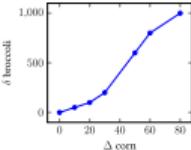


Figure 2: The broccoli coefficient δ broccoli in relation to the corn coefficient Δ corn.

Etiam pede massa, dapibus vitae, rhoncus in, placerat posuere, odio. Vestibulum luctus commodo laus. Morbi laces dui, tempor nisl, et, egestas, condimentum at, tortor. Phasellus aliquet odio ac lacinia tempore facilisis. Praesent sed sem. Praesent iaculis. Cras rhoncus tellus sed justo ullamcorper sagittis. Donec quis orci. Sed ut tortor quis tellus euismod tincidunt. Suspendisse congue nisi eu elit. Aliquam tortor diam, tempus id, tristique eget, sodales vel, nulla. Praesent tellus mi, condimentum sed, viverra at, consectetur quis, lectus. In auctor vehicula orci. Sed pede sapien, euismod in, suscipit in, pharetra placerat, metus. Vivamus commodo dui non odio.

Etiam suscipit aliquam arcu. Aliquam sit amet est ac purus bibendum congue. Sed in eros. Morbi non orci. Pellentesque mattis lacinia elit. Fusce molestie velit in ligula. Nullam et orci vitae nibh vulputate auctor. Aliquam eget purus. Nulla auctor wisi sed ipsum. Morbi portitor tellus ac enim. Fusce ornare. Proin ipsum enim, tincidunt in, ornare venenatis, molestie a, augue. Donec vel pede in lacinia sagittis porta. Sed hendrerit ipsum quis nisi. Suspendisse quis massa ac nibh pretium cursum. Sed sodales. Nam eu neque quis pede dignissim ornare. Maecenas eu purus ac urna tincidunt.

A comparison of the two methods is shown in Figure 1. Vivamus eu tellus sed tellus consequatur suscipit. Nam orci orci, interdum quis, porttitor nec, ultricies vita, erat. Donec risus turpis, luctus sit amet, interdum quis, porttitor nec, ipsum. Suspendisse condimentum, tortor at egestas posuere, neque metus tempore orci, et tincidunt urna mince a purus. Sed facilisis blandit tellus. Nunc risus sem, suscipit nec, eleifend quis, cursus quis, libero. Curabitur et dolor. Sed

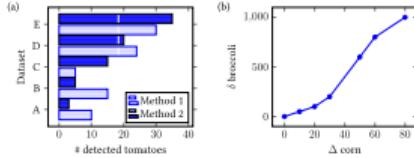


Figure 3: (a) The number of tomatoes detected by either of the two methods across the five main datasets. (b) The broccoli coefficient δ brocoli in relation to the corn coefficient Δ corn.

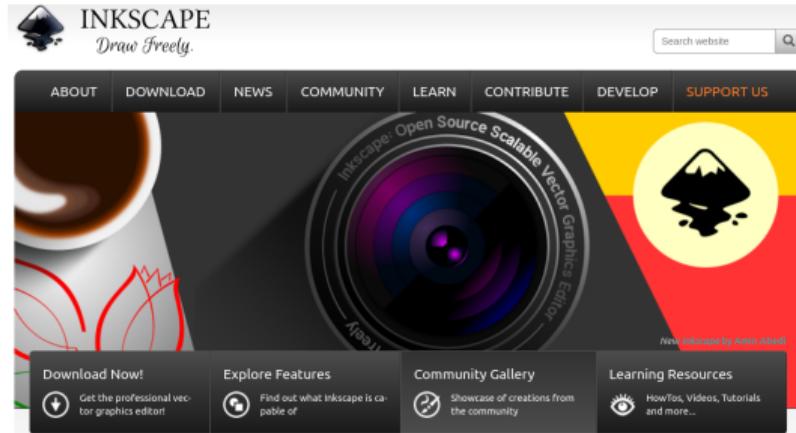
Etiam pede massa, dapibus vitae, rhoncus in, placerat posuere, odio. Vestibulum luctus commodo laus. Morbi laces dui, tempor nisl, et, egestas, condimentum at, tortor. Phasellus aliquet odio ac lacinia tempore facilisis. Praesent sed sem. Praesent iaculis. Cras rhoncus tellus sed justo ullamcorper sagittis. Donec quis orci. Sed ut tortor quis tellus euismod tincidunt. Suspendisse congue nisi eu elit. Aliquam tortor diam, tempus id, tristique eget, sodales vel, nulla. Praesent tellus mi, condimentum sed, viverra at, consectetur quis, lectus. In auctor vehicula orci. Sed pede sapien, euismod in, suscipit in, pharetra placerat, metus. Vivamus commodo dui non odio. Donec et felis.

Etiam suscipit aliquam arcu. Aliquam sit amet est ac purus bibendum congue. Sed in eros. Morbi non orci. Pellentesque mattis lacinia elit. Fusce molestie velit in ligula. Nullam et orci vitae nibh vulputate auctor. Aliquam eget purus. Nulla auctor wisi sed ipsum. Morbi portitor tellus ac enim. Fusce ornare. Proin ipsum enim, tincidunt in, ornare venenatis, molestie a, augue. Donec vel pede in lacinia sagittis porta. Sed hendrerit ipsum quis nisi. Suspendisse quis massa ac nibh pretium cursum. Sed sodales. Nam eu neque quis pede dignissim ornare. Maecenas eu purus ac urna tincidunt.

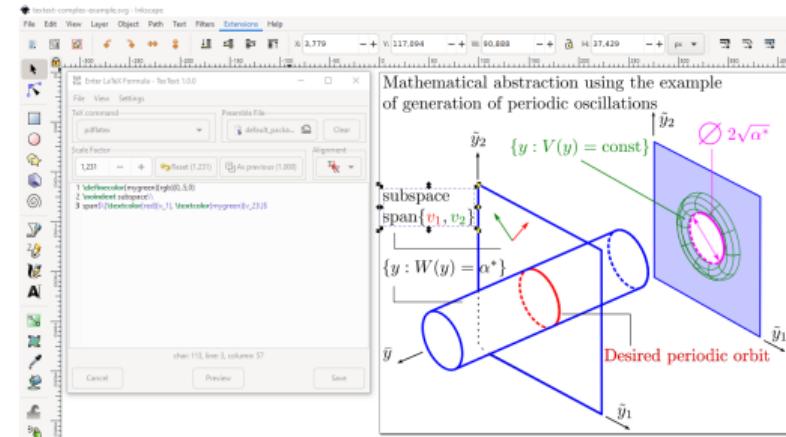
A comparison of the two methods is shown in Figure 3(a). Vivamus eu tellus sed tellus consequatur suscipit. Nam orci orci, interdum quis, porttitor nec, ultricies vita, erat. Donec risus turpis, luctus sit amet, interdum quis, porttitor nec, ipsum. Suspendisse condimentum, tortor at egestas posuere, neque metus tempore orci, et tincidunt urna mince a purus. Sed facilisis blandit tellus. Nunc risus sem, suscipit nec, eleifend quis, cursus quis, libero. Curabitur et dolor. Sed

et, pede. Donec wisi sapien, feugiat in, porttitor nec, blandit tincidunt adipiscing, metus. Donec vel nibh ut felis consectetur laoreet. Donec pede. Sed id quam id wisi laoreet suscipit. Nulla luctus dolore, aliquam ac, fringilla eget, mollis ut, orci. In pellentesque justo in

How to create illustrations?



<https://inkscape.org/>



<https://texttext.github.io/texttext/>

- ▶ **Vector graphics** program, i.e., Inkscape
 - ▶ Typeset L^AT_EX inside Inkscape using **TexText**

Follow good scientific practice

- ▶ Your text should be **your own exposition** and explain things in your words
- ▶ **Do not copy** sentences 1:1 from your sources (unusual in natural science)
- ▶ Be inspired by the papers you read, adopt good writing styles
- ▶ Use tools like www.grammarly.com for finding and fixing typos
- ▶ **Do not use un-edited GPT outputs unless explicitly permitted!**

Cite everything relevant

- ▶ Whenever stating a fact that is known, add the corresponding citations
- ▶ Make sure all related work is cited appropriately
- ▶ Citations are added before punctuation marks, e.g.: “.. as illustrated in [15].”
- ▶ Use \LaTeX in combination with **Bibtex** to manage your citations and bibliography
https://www.overleaf.com/learn/latex/Bibtex_bibliography_styles
- ▶ Use the `cite` package to format the bibliography alphabetically

2.3

Reviewing

What is a review?

- ▶ Reviews judge if a paper gets accepted
- ▶ 3-5 reviews / paper
- ▶ Area chairs / associate editors make final decision based on reviews
- ▶ Top conferences/journals have acceptance rates <25%
- ▶ Often the authors and reviewers don't know each other (double blind)
- ▶ Sometimes the reviewers can see the author's names (single blind)

Why not accept everything?

Papers can have a negative impact:

- ▶ Wrong or fraudulent results mislead the field and damage the reputation of the conference
- ▶ Misleading evaluation makes it hard to compare with, kills follow-up
- ▶ Creates bad precedent (weak paper X got in, so this one should too)
- ▶ Fatigue/overload of too many papers, wastes everyone's time

Why should I care?

- ▶ Understanding reviewing → **better reading and writing!**
 - ▶ Critical thinking
 - ▶ Better notes when reading
 - ▶ Concise, unambiguous writing
 - ▶ Better structuring
- ▶ You may be invited to review in the future

Read the review guidelines!

Example: TMLR acceptance criteria

- ▶ Are the claims made in the submission supported by accurate, convincing and clear evidence?
- ▶ Would some individuals in TMLR's audience be interested in the findings of this paper?

Read the review guidelines!

Example: CVPR/ICCV acceptance criteria

Any paper that, with CVPR/ICCV community standards,

- ▶ presents sufficient knowledge advancement that is well grounded
- ▶ is of sufficient interest to some CVPR/ICCV audiences who could benefit from it

Key points

- ▶ Provide **feedback** to the authors prior to publication, including:
Language, clarity, rigor, references, experiments (and novelty)
- ▶ Provide a **recommendation** to the AC with clear reasoning
- ▶ Ultimate goal to improve the manuscript → **concrete suggestions**
- ▶ Reviews are **objective** and state both **pros and cons**
- ▶ **"Review"** the review from the perspective of authors and AC

The general review structure



Summary:

Describe the key ideas, experiments, and their significance (preferably in 5-7 sentences).



Strengths:

Consider the aspects of key ideas, experimental or theoretical validation, writing quality, and data contribution (if relevant). Explain clearly why these aspects of the paper are valuable.



Weaknesses:

Consider the aspects of key ideas, experimental or theoretical validation, writing quality, and data contribution (if relevant). Explain clearly why these are weak aspects of the paper.



Rating and Justification:

Provide detailed justification of your rating. It should involve how you weigh the strengths and weaknesses of the paper.



Additional comments:

Minor suggestions, questions, corrections, etc. that can help the authors improve the paper, if any.

Different papers typically need different results

- ▶ Established problem, plausible idea → **benchmark results**
- ▶ Weird, complex, and/or implausible → **extraordinary results** (which need to be scrutinized carefully)
- ▶ Potentially transformative idea → **basic proof-of-concept**
- ▶ Position piece or theory paper → **no experiments**

Where to read reviews

The screenshot shows the OpenReview.net homepage. At the top, there is a navigation bar with the site's logo, a search bar, and a login link. Below the navigation bar, a banner displays the text: "Open Peer Review. Open Publishing. Open Access. Open Discussion. Open Recommendations. Open Directory. Open API. Open Source." The main content area is divided into two sections: "Active Venues" on the left and "Open for Submissions" on the right.

Active Venues	Open for Submissions
TMLR	RSS 2023 Workshop Robotic Assembly Due 07 Jun 2023, 13:59 Central European Summer Time
ACM ICMI 2022 Workshop MCI	ACMMM 2023 Track Brave New Ideas Due 08 Jun 2023, 01:59 Central European Summer Time
AI4Science 2022 Internal PBS	ACMMM 2023 Track Open Source Due 08 Jun 2023, 02:00 Central European Summer Time
EMNLP 2022 Workshop LOUHI	EWRL 2023 Workshop Due 08 Jun 2023, 13:59 Central European Summer Time
DH 2022 Workshop BD	KDD 2023 Workshop DSAI4Sports Due 08 Jun 2023, 14:59 Central European Summer Time
VU Amsterdam 2023 PrIns	ICML 2023 Workshop SODS Due 09 Jun 2023, 01:59 Central European Summer Time
ACM IUI 2023 Workshop ITAH	CoRL 2023 Conference Due 09 Jun 2023, 08:59 Central European Summer Time
Open Life Science 2023 Cohort 7	IWAI 2023 Workshop Due 10 Jun 2023, 14:00 Central European Summer Time
MASC-SLL 2023 Colloquium	
HRI 2023 Workshop VAM-HRI	
JSYS 2023 March Papers	
SupaeroSDD 2023 Workshop	

<https://openreview.net>

- ▶ Reviews **publicly available** for ICLR, TMLR, NeurIPS, CoRL...
- ▶ Search like google scholar

Thank You!

<https://kashyap7x.github.io>