

# 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 interface. The search bar at the top contains the text "end-to-end autonomous driving" and is highlighted with a red box. Below the search bar, the results are listed. Three results are visible, each with a red box around its PDF link:

- Result 1: "Multimodal end-to-end autonomous driving" by Y. Xiao, F. Codreanu, A. Gurnam. PDF link: [PDF] arxiv.org
- Result 2: "Multi-modal fusion transformer for end-to-end autonomous driving" by A. Prakash, K. Chitta, A. Geiger. PDF link: [PDF] thecvf.com
- Result 3: "End to end learning for self-driving cars" by M. Bojarski, D. Del Testa, D. Dworakowski. PDF link: [PDF] arxiv.org

`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-

ject detection (see [58]). To overcome this limitation, 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 possible positions!

**News:** I was recently selected as an *RSJ* reviewer 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 *nuPlan 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 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. *Murtal Habert*. During this time, he was also an intern at *NYDA* working with Dr. *Jose M. Alvarez*. He is currently a PhD student in the *Autonomous Vision Group* at the University of Tübingen, Germany, supervised by Prof. *Andreas Geiger*.

[CV](#) [Mail](#) [Scholar](#) [Twitter](#) [LinkedIn](#) [Facebook](#) [Mastodon](#) [BibHub](#) [YouTube](#)



### Publications



**PlanT: Explainable Planning Transformers via Object-Level Representations**  
Katriin Rezc, **Kashyap Chitta**, Oshri Bogdan Meroz, Sophia Koenke, Zeynep Akinci, Andreas Geiger  
Conference on Robot Learning (CoRL), 2022  
[Abs](#) / [Paper](#) / [Supplementary](#) / [Video](#) / [Code](#) / [BibTex](#)



**KING: Generalizing Safety Critical Driving Scenarios for Robust Imitation via Kinematic Gradients**  
Niklas Haaselmayer, Katriin Rezc, **Kashyap Chitta**, Aparithi 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, Zehao Yu, Katriin Rezc, Andreas Geiger  
Transactions on Pattern Analysis and Machine Intelligence (TPAMI), 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**  
Marissa Wirt, **Kashyap Chitta**, Yash Sharma, Wladimir Brändt, Matthias Berghy, 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, Eshad Chn-Bar, Andreas Geiger  
International Conference on Intelligent Robots and Systems (IROS), 2020  
[Abs](#) / [Paper](#) / [Video](#) / [BibTex](#)



**Scalable Active Learning for Object Detection**  
Elmar Haussmann, Michelle Fendt, **Kashyap Chitta**, Jan Isenack, Hanson Xu, Dorena Roy, Akshita Mittal, Nicolas Koumchatzky, Clemens Farabet, Jose Alvarez  
Intelligent Vehicles Symposium (IV), 2020  
[Abs](#) / [Paper](#) / [BibTex](#)



**Learning Situational Driving**  
Eshad Chn-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, Eshad Chn-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

CONNECTED PAPERS

multi-modal fusion transformer for end-to-end autonomous driving

Share Follow About Pricing Sponsors Log in

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

Prior works Derivative works Sponsored by Azure

Search... Expand

Origin paper

**Multi-Modal Fusion Transformer for End-to-End Autonomous Driving**  
Aditya Prakash, Kashyap Chitta, Andreas Geiger 2021

TransFuser: Imitation with Transformer-Based Sensor Fusion for Autonomous Driving  
Kashyap Chitta, Aditya Prakash, Bernhard... 2022

NEAT: Neural Attention Fields for End-to-End Autonomous Driving  
Kashyap Chitta, Aditya Prakash, Andreas Geiger 2021

Learning by Watching  
Jimuyang Zhang, Eshed Ohn-Bar 2021

Safety-Enhanced Autonomous Driving Using Interpretable Sensor Fusion Transformer  
Hao-Chiang Shao, Letian Wang, Ruobing Chen,... 2022

Multi-View Fusion of Sensor Data for Improved Perception and Prediction in Autonomous...  
Sudeep Fadadu, Shreyash Pandey, Darshan... 2020

End-to-end Contextual Perception and

Created on May 21, 2023

2019 2023

Multi-Modal Fusion Transformer for End-to-End Autonomous Driving  
Aditya Prakash, Kashyap Chitta, Andreas Geiger  
2021, Computer Vision and Pattern Recognition ...

181 Citations Save

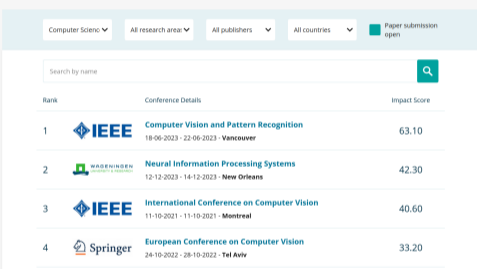
Open in: PDF BibTeX RSS DOI GitHub X





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



# 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/>



Rank	Conference Details	Impact Score
1	 <b>Computer Vision and Pattern Recognition</b> 18-06-2023 - 22-06-2023 - Vancouver	63.10
2	 <b>Neural Information Processing Systems</b> 12-12-2023 - 14-12-2023 - New Orleans	42.30
3	 <b>International Conference on Computer Vision</b> 11-10-2021 - 11-10-2021 - Montreal	40.60
4	 <b>European Conference on Computer Vision</b> 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

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.

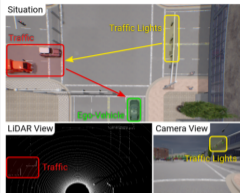
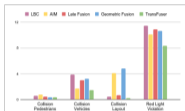


Figure 1: **Illustration.** Consider an intersection with oncoming traffic from the left. To safely navigate the intersection, the ego-vehicle (green) must capture the global context of the scene involving the interaction between the traffic light (yellow) and the vehicles (red). However, the traffic 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



(b) **Infractions.** We report the mean value of the total infractions incurred by each model over the 9 evaluation runs in the Town05 Short setting.

- ▶ 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 Karkus<sup>1</sup>, Boris Ivanovic<sup>1</sup>, Shie Mannor<sup>1,2</sup>, Marco Pavone<sup>1,3</sup>

<sup>1</sup>NVIDIA Research, <sup>2</sup>Technion, <sup>3</sup>Stanford 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](http://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,  $\frac{\delta \mathcal{L}_{\text{plan}}}{\delta \theta} = \frac{\delta \mathcal{L}_{\text{CE}}}{\delta p_n} \frac{\delta p_n}{\delta C} \frac{\delta C}{\delta C_{\text{coll}}} \frac{\delta C_{\text{coll}}}{\delta s_a} \frac{\delta s_a}{\delta \theta}$ , 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

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

where  $C$  denotes the cost function,  $f_d$  the dynamics,  $s^{\text{init}}$  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  $u_{\text{plan}}$  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](#),<sup>[1]</sup> 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<sup>\*†1</sup>, Boris Ivanovic<sup>\*1</sup>, Punarjay Chakravarty<sup>2</sup>, and  
Marco Pavone<sup>1</sup>

<sup>1</sup> Autonomous Systems Lab, Stanford University  
{timsal, borisi, pavone}@stanford.edu

<sup>2</sup> Ford Greenfield Labs  
pchakra5@ford.com

# Use project pages

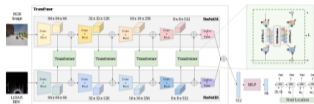
- ▶ Often contain:
  - ▶ Talks and **slides**
  - ▶ Narrated **videos**
  - ▶ Supplementary materials
  - ▶ Source code (e.g., github)
  - ▶ Additional resources (e.g., blog)
- ▶ Use these resources to quickly get a high-level understanding of a paper
- ▶ For more tips on reading, see **Jia-Bin Huang's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

### Abstract

The global representations from complementary sensors are integrated for autonomous driving. Geometry-based sensor fusion has shown great promise for perception tasks such as object detection and scene forecasting. However, for the end-to-end 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 efficiently fusing representations in end-to-end driving tasks. To this end, we demonstrate that decision 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 resulting from multiple directions at uncontrolled intersections. Therefore, we propose Transformer, a novel Multi-modal Fusion Transformer, to integrate scene and LiDAR representations using attention. We experimentally validate the efficacy of our approach in other settings involving complex scenarios using the CARLA urban driving simulator. Our approach achieves state-of-the-art driving performance while reducing collisions by 35% compared to geometry-based fusion.

### Video



### Generalization to New Town







# Which tool can I use to write?

3DV #	3DV 2021 Submission #, CONFIDENTIAL REVIEW COPY, DO NOT DISTRIBUTE.	3DV #
900		904
901		905
902		906
903	<b>Your Title</b>	907
904	3DV Seminar Report	908
905		909
906		910
907	Firstname Lastname    Firstname Lastname	911
908		912
909		913
910		914
911	<b>1. Introduction</b>	915
912	This is a test citation [1].	916
913		917
914		918
915	<b>2. Methods</b>	919
916		920
917	<b>3. Evaluation</b>	921
918		922
919	<b>4. Discussion</b>	923
920		924
921		925
922	<b>References</b>	926
923	[1] Michael Numentor and Andrew Grigor. Graphs: Representing scenes as compositional generative neural feature fields. In Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), 2020. 1	927
924		928
925		929
926		930
927		931
928		932
929		933
930		934
931		935
932		936
933		937
934		938
935		939
936		940
937		941
938		942
939		943
940		944
941		945
942		946
943		947
944		948
945		949
946		950
947		951
948		952
949		953
950		954
951		955
952		956
953		957
954		958
955		959
956		960
957		961
958		962
959		963
960		964
961		965
962		966
963		967
964		968
965		969
966		970
967		971
968		972
969		973
970		974
971		975
972		976
973		977
974		978
975		979
976		980
977		981
978		982
979		983
980		984
981		985
982		986
983		987
984		988
985		989
986		990
987		991
988		992
989		993
990		994
991		995
992		996
993		997
994		998
995		999
996		1000
997		1001
998		1002
999		1003
1000		1004
1001		1005
1002		1006
1003		1007
1004		1008
1005		1009
1006		1010
1007		1011
1008		1012
1009		1013
1010		1014
1011		1015
1012		1016
1013		1017
1014		1018
1015		1019
1016		1020
1017		1021
1018		1022
1019		1023
1020		1024
1021		1025
1022		1026
1023		1027
1024		1028
1025		1029
1026		1030
1027		1031
1028		1032
1029		1033
1030		1034
1031		1035
1032		1036
1033		1037
1034		1038
1035		1039
1036		1040
1037		1041
1038		1042
1039		1043
1040		1044
1041		1045
1042		1046
1043		1047
1044		1048
1045		1049
1046		1050
1047		1051
1048		1052
1049		1053
1050		1054
1051		1055
1052		1056
1053		1057
1054		1058
1055		1059
1056		1060
1057		1061
1058		1062
1059		1063
1060		1064
1061		1065
1062		1066
1063		1067
1064		1068
1065		1069
1066		1070
1067		1071
1068		1072
1069		1073
1070		1074
1071		1075
1072		1076
1073		1077
1074		1078
1075		1079
1076		1080
1077		1081
1078		1082
1079		1083
1080		1084
1081		1085
1082		1086
1083		1087
1084		1088
1085		1089
1086		1090
1087		1091
1088		1092
1089		1093
1090		1094
1091		1095
1092		1096
1093		1097
1094		1098
1095		1099
1096		1100
1097		1101
1098		1102
1099		1103
1100		1104
1101		1105
1102		1106
1103		1107
1104		1108
1105		1109
1106		1110
1107		1111
1108		1112
1109		1113
1110		1114
1111		1115
1112		1116
1113		1117
1114		1118
1115		1119
1116		1120
1117		1121
1118		1122
1119		1123
1120		1124
1121		1125
1122		1126
1123		1127
1124		1128
1125		1129
1126		1130
1127		1131
1128		1132
1129		1133
1130		1134
1131		1135
1132		1136
1133		1137
1134		1138
1135		1139
1136		1140
1137		1141
1138		1142
1139		1143
1140		1144
1141		1145
1142		1146
1143		1147
1144		1148
1145		1149
1146		1150
1147		1151
1148		1152
1149		1153
1150		1154
1151		1155
1152		1156
1153		1157
1154		1158
1155		1159
1156		1160
1157		1161
1158		1162
1159		1163
1160		1164
1161		1165
1162		1166
1163		1167
1164		1168
1165		1169
1166		1170
1167		1171
1168		1172
1169		1173
1170		1174
1171		1175
1172		1176
1173		1177
1174		1178
1175		1179
1176		1180
1177		1181
1178		1182
1179		1183
1180		1184
1181		1185
1182		1186
1183		1187
1184		1188
1185		1189
1186		1190
1187		1191
1188		1192
1189		1193
1190		1194
1191		1195
1192		1196
1193		1197
1194		1198
1195		1199
1196		1200
1197		1201
1198		1202
1199		1203
1200		1204
1201		1205
1202		1206
1203		1207
1204		1208
1205		1209
1206		1210
1207		1211
1208		1212
1209		1213
1210		1214
1211		1215
1212		1216
1213		1217
1214		1218
1215		1219
1216		1220
1217		1221
1218		1222
1219		1223
1220		1224
1221		1225
1222		1226
1223		1227
1224		1228
1225		1229
1226		1230
1227		1231
1228		1232
1229		1233
1230		1234
1231		1235
1232		1236
1233		1237
1234		1238
1235		1239
1236		1240
1237		1241
1238		1242
1239		1243
1240		1244
1241		1245
1242		1246
1243		1247
1244		1248
1245		1249
1246		1250
1247		1251
1248		1252
1249		1253
1250		1254
1251		1255
1252		1256
1253		1257
1254		1258
1255		1259
1256		1260
1257		1261
1258		1262
1259		1263
1260		1264
1261		1265
1262		1266
1263		1267
1264		1268
1265		1269
1266		1270
1267		1271
1268		1272
1269		1273
1270		1274
1271		1275
1272		1276
1273		1277
1274		1278
1275		1279
1276		1280
1277		1281
1278		1282
1279		1283
1280		1284
1281		1285
1282		1286
1283		1287
1284		1288
1285		1289
1286		1290
1287		1291
1288		1292
1289		1293
1290		1294
1291		1295
1292		1296
1293		1297
1294		1298
1295		1299
1296		1300
1297		1301
1298		1302
1299		1303
1300		1304
1301		1305
1302		1306
1303		1307
1304		1308
1305		1309
1306		1310
1307		1311
1308		1312
1309		1313
1310		1314
1311		1315
1312		1316
1313		1317
1314		1318
1315		1319
1316		1320
1317		1321
1318		1322
1319		1323
1320		1324
1321		1325
1322		1326
1323		1327
1324		1328
1325		1329
1326		1330
1327		1331
1328		1332
1329		1333
1330		1334
1331		1335
1332		1336
1333		1337
1334		1338
1335		1339
1336		1340
1337		1341
1338		1342
1339		1343
1340		1344
1341		1345
1342		1346
1343		1347
1344		1348
1345		1349
1346		1350
1347		1351
1348		1352
1349		1353
1350		1354
1351		1355
1352		1356
1353		1357
1354		1358
1355		1359
1356		1360
1357		1361
1358		1362
1359		1363
1360		1364
1361		1365
1362		1366
1363		1367
1364		1368
1365		1369
1366		1370
1367		1371
1368		1372
1369		1373
1370		1374
1371		1375
1372		1376
1373		1377
1374		

# Overleaf

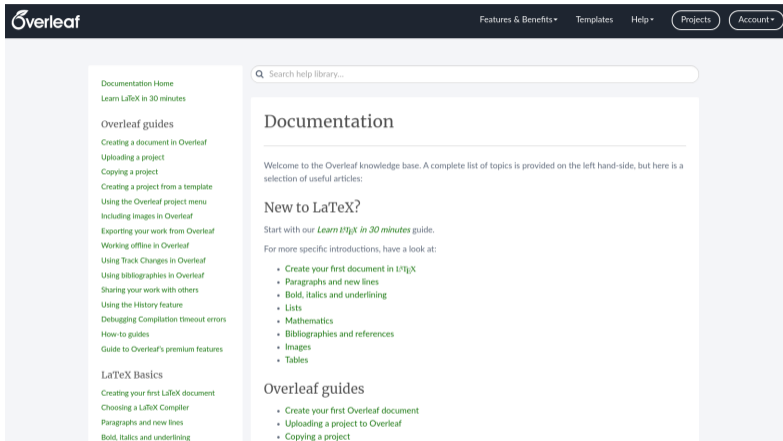
The screenshot displays the Overleaf online LaTeX editor interface. The top bar includes navigation and utility icons such as Menu, Review, Share, Submit, History, and Chat. The main workspace is divided into two panes:

- Source Pane (Left):** Shows the LaTeX source code for a document titled "University of Tübingen 3D Vision Report Template". The code includes package loading, document structure commands, and content for sections like Introduction, Methods, Evaluation, and Discussion. A file outline on the left side of this pane lists the document's structure.
- PDF Pane (Right):** Shows the rendered PDF output. The document title is "3DV Seminar Report" and the author information is "Firstname Lastname Firstname Lastname". The content includes a table of contents, section headers (1. Introduction, 2. Methods, 3. Evaluation, 4. Discussion), and a reference list.

[www.overleaf.com](http://www.overleaf.com)

- Online, no installation, good for beginners

# How can I learn $\text{\LaTeX}$ ?



The screenshot shows the Overleaf documentation page. At the top, there is a dark navigation bar with the Overleaf logo on the left and links for 'Features & Benefits', 'Templates', 'Help', 'Projects', and 'Account' on the right. Below the navigation bar is a search bar with the placeholder text 'Search help library...'. The main content area is divided into a left sidebar and a main content column. The sidebar contains a list of links under the following categories: 'Documentation Home', 'Learn LaTeX in 30 minutes', 'Overleaf guides', and 'LaTeX Basics'. The main content column has a heading 'Documentation' followed by a welcome message. Below that is a section titled 'New to LaTeX?' with a list of articles. At the bottom of the main content column is a section titled 'Overleaf guides' with a list of articles.

Overleaf

Features & Benefits Templates Help Projects Account

Search help library...

Documentation

Welcome to the Overleaf knowledge base. A complete list of topics is provided on the left hand-side, but here is a selection of useful articles:

### New to LaTeX?

Start with our *Learn  $\text{\LaTeX}$  in 30 minutes* guide.

For more specific introductions, have a look at:

- Create your first document in  $\text{\LaTeX}$
- Paragraphs and new lines
- Bold, italics and underlining
- Lists
- Mathematics
- Bibliographies and references
- Images
- Tables

### Overleaf guides

- Create your first Overleaf document
- Uploading a project to Overleaf
- Copying a project

[www.overleaf.com/learn](http://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

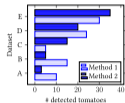


Figure 1: The number of tomatoes detected by either of the two methods across the five main datasets.

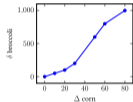


Figure 2: The broccoli coefficient  $\delta$  broccoli in relation to the corn coefficient  $\Delta$  corn.

Etiam pede massa, dapibus vitae, rhoncus in, placerat posuere, odio. Vestibulum luctus commodo lacus. Morbi lacus dui, tempor sed, euismod eget, condimentum at, tortor. Phasellus aliquet odio ac lacus tempor faucibus. 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, phanetra 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 lectus elit. Fusce molestie velit in ligula. Nullam et orci vitae nibh vulpate 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 lacus sagittis porta. Sed hendrerit ipsum quis nisi. Suspendisse quis massa ac nibh pretium cursus. Sed sodales. Nam eu neque quis pede dignissim ornare. Maecenas eu purus ac urna tincidunt.

A comparison of the two methods (see Figure 1). Vivamus eu tellus sed tellus consequat suscipit. Nam orci orci, donec wisi, ultricies vitae, erat. Donec risus turpis, luctus sit amet, interdum quis, pede sem ipsum. Suspendisse condimentum, tortor at egestas posuere, neque metus tempor orci, et tincidunt urna nunc a purus. Sed facilisis blandit tellus. Nunc risus sem, suscipit nec, eleifend quis, cursus quis, libero. Curabitur et dolor. Sed



3

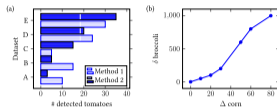


Figure 3: (a) The number of tomatoes detected by either of the two methods across the five main datasets. (b) The broccoli coefficient  $\delta$  broccoli in relation to the corn coefficient  $\Delta$  corn.

Etiam pede massa, dapibus vitae, rhoncus in, placerat posuere, odio. Vestibulum luctus commodo lacus. Morbi lacus dui, tempor sed, euismod eget, condimentum at, tortor. Phasellus aliquet odio ac lacus tempor faucibus. 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, phanetra 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 lectus elit. Fusce molestie velit in ligula. Nullam et orci vitae nibh vulpate 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 lacus sagittis porta. Sed hendrerit ipsum quis nisi. Suspendisse quis massa ac nibh pretium cursus. Sed sodales. Nam eu neque quis pede dignissim ornare. Maecenas eu purus ac urna tincidunt.

A comparison of the two methods (see Figure 1). Vivamus eu tellus sed tellus consequat suscipit. Nam orci orci, donec wisi, ultricies vitae, erat. Donec risus turpis, luctus sit amet, interdum quis, pede sem ipsum. Suspendisse condimentum, tortor at egestas posuere, neque metus tempor orci, et tincidunt urna nunc a purus. Sed facilisis blandit tellus. Nunc risus sem, suscipit nec, eleifend quis, cursus quis, libero. Curabitur et dolor. Sed vitae sem. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Maecenas ante. Donec wisi, ultricies vitae, erat. Donec risus turpis, luctus sit amet, interdum quis, pede sem ipsum. Suspendisse condimentum, tortor at egestas posuere, neque metus tempor orci, et tincidunt urna nunc a purus. Sed facilisis blandit tellus. Nunc risus sem, suscipit nec, eleifend quis, cursus quis, libero. Curabitur et dolor. Sed

Duis aliquet dui in eros. Suspendisse condimentum, in, accumsan non, enim. Aliquam erat volutpat. 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, phanetra placerat, metus. Vivamus commodo dui non odio. Donec et felis.

Donec vel pede in lacus sagittis porta. Sed hendrerit ipsum quis nisi. Suspendisse quis massa ac nibh pretium cursus. Sed sodales. Nam eu neque quis pede dignissim ornare. Maecenas eu purus ac urna tincidunt.

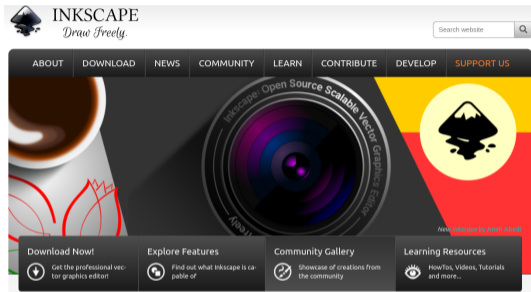
Donec vel pede in lacus sagittis porta. Sed hendrerit ipsum quis nisi. Suspendisse quis massa ac nibh pretium cursus. Sed sodales. Nam eu neque quis pede dignissim ornare. Maecenas eu purus ac urna tincidunt.



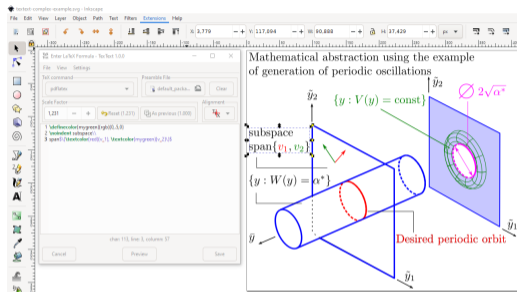
3



# How to create illustrations?



<https://inkscape.org/>



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

- ▶ **Vector graphics** program, i.e., Inkscape
- ▶ Typeset  $\\text{\\LaTeX}$  inside Inkscape using **TextText**

## 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](http://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  $\text{\LaTeX}$  in combination with **Bibtex** to manage your citations and bibliography  
[https://www.overleaf.com/learn/latex/Bibtex\\_bibliography\\_styles](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, un-ambiguous 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.



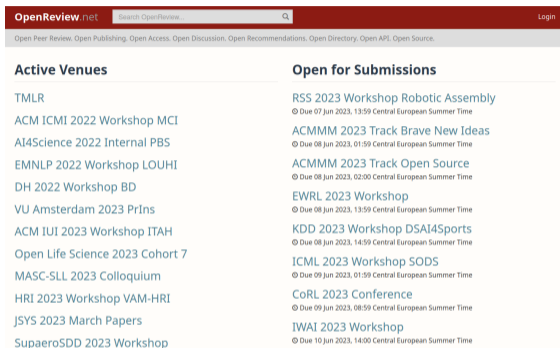
## Additional comments:

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 OpenReview.net logo, a search bar, and a login link. Below the navigation bar, there are two main sections: "Active Venues" and "Open for Submissions".

**Active Venues**

- TMLR
- ACM ICMI 2022 Workshop MCI
- AI4Science 2022 Internal PBS
- EMNLP 2022 Workshop LOUHI
- DH 2022 Workshop BD
- VU Amsterdam 2023 PrIns
- ACM IUI 2023 Workshop ITAH
- Open Life Science 2023 Cohort 7
- MASC-SLL 2023 Colloquium
- HRI 2023 Workshop VAM-HRI
- JSYS 2023 March Papers
- SupaeroSDD 2023 Workshop

**Open for Submissions**

- RSS 2023 Workshop Robotic Assembly  
⌚ Due 07 Jun 2023, 13:59 Central European Summer Time
- ACMMM 2023 Track Brave New Ideas  
⌚ Due 08 Jun 2023, 01:59 Central European Summer Time
- ACMMM 2023 Track Open Source  
⌚ Due 08 Jun 2023, 02:00 Central European Summer Time
- EWRL 2023 Workshop  
⌚ Due 08 Jun 2023, 13:59 Central European Summer Time
- KDD 2023 Workshop DSAI4Sports  
⌚ Due 08 Jun 2023, 14:59 Central European Summer Time
- ICML 2023 Workshop SODS  
⌚ Due 09 Jun 2023, 01:59 Central European Summer Time
- CoRL 2023 Conference  
⌚ Due 09 Jun 2023, 08:59 Central European Summer Time
- IWAI 2023 Workshop  
⌚ Due 10 Jun 2023, 14:00 Central European Summer Time

<https://openreview.net>

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

# Thank You!

<https://kashyap7x.github.io>