

# 唐涛

(+86) 15755623037 | [trent.tangtao@gmail.com](mailto:trent.tangtao@gmail.com) | [tangtao.github.io/](https://tangtao.github.io/)

## 教育背景

### 中山大学 | 博士学位

2021.09—2026.06 (预计)

智能工程学院, 中山大学 HCP 实验室, 导师: 梁小丹教授, 2026 年应届生。

主要研究方向: 计算机视觉, 自动驾驶。发表顶会/刊论文 14 篇, 其中一作 5 篇 (含共一), 谷歌学术引用 1000+。

### 大连理工大学 | 学士学位

2017.09—2021.06

软件学院, 专业排名第二, 获国家奖学金、辽宁省优秀毕业生、玉兰奖学金等奖项。

## 实习经历

### 理想汽车 | 自动驾驶实验室 | 世界模型实习生

2023.12—至今

- **自动驾驶场景重建**: 自动驾驶传感器仿真以及数据生成是数据闭环的重要一环: 1. 我们提出了第一个基于神经渲染的 LiDAR 仿真生成器 LiDAR-NeRF[4], 其可以端到端生成 LiDAR 点云的新视角。2. 并且我们将其拓展到多模态上, 提出了 AlignMiF[3] 解决了多模态的不一致问题, 实现 LiDAR-图像联合多模态统一神经场的可微分渲染。3. 随着 3DGS 的发展, 我们还探索了 LiDAR-GS[7], 将 NeRF 的表达升级为 GS 实现了更快更好的 LiDAR 渲染。4. 对于机器人场景, 我们同样利用 GS 进行了场景仿真, 提出了 RoboPearls[2] 开发了语义增强的动态场景重建, 同时结合 LLM 支持用户用自然语言进行场景编辑仿真, 有效增强了下游机器人操纵性能。
- **自动驾驶场景生成**: 利用生成模型生成自驾场景和数据已经非常流行: 1. 参与设计了 Delphi[6], 一种基于扩散模型的长视频生成方法, 通过跨多视角的共享噪声建模机制来增加空间一致性, 并通过特征对齐模块实现精确的可控性和时间一致性。2. 我们提出了第一个端到端 LiDAR-图像联合生成框架 OmniGen[1], 我们首先提出了一个多模态的自编码器, 将多模态输入都编码到统一的 BEV 空间, 然后利用扩散模型生成对应的 BEV 特征, 再由自编码器解码出多模态的生成结果。3. 针对自驾场景检索: 已有的图像检索缺乏场景全局特征表示, 很难和复杂的场景描述对应, 我们提出了 BEV-TSR[9] 第一种多模态 BEV 检索方法, 在 BEV 空间检索相应的场景, 同时利用 LLM 并结合知识图谱中的半结构化信息, 提高语义特征提取能力, 实现了高性能的文本-场景检索。
- **端到端算法**: 1. 基于 VLM 的端到端规划算法: 我们提出了一个推理决策一致性的 RDA-Drive[10], 现有大语言模型与自动驾驶结合的工作都忽略了因果关系中的不一致性问题, 我们应用排序对齐损失来约束 CoT 和规划结果之间的推理-决策一致性, 从而实现了最先进的规划性能。2. 端到端 3D 跟踪: 我们提出了一个简单鲁棒的端到端 3DMOT 模型 S2-Track[8], 霸榜 nuScenes 至今。对于 query 初始化, 我们利用 2D 目标位置和深度信息估计出初始 3D 位置, 对于 query 优化, 我们引入了一个不确定性感知的概率解码器, 将注意力建模为高斯分布而不是确定性的输出, 以量化预测跟踪的不确定性, 对于 query 训练, 我们向真实框添加噪声后去噪, 增强了目标框匹配的稳定性。

### 阿里巴巴 | 达摩院 | 自动驾驶实验室 | 感知算法实习生

2022.03—2023.12

- **自动驾驶感知算法**: 研发鲁棒可靠的感知算法。针对 3D 检测: 1. 我们提出了 RobustBenchmark[12], 收集了一系列 BEV 感知中噪声数据的真实案例, 并系统地制定了一个鲁棒性基准的工具包来模拟这些案例, 而且在广泛采用 NuScenes 和 Waymo 数据集上建立了两个新的鲁棒性基准来展示其有效性, 然后全面评估了流行的 BEV 感知的融合方法; 2. 参与了 BEVFusion[15] 框架的研究, 和之前的方法不同的是雷达点云的处理和图像的处理是独立进行的, 利用单独的网络进行编码, 投射到统一的 BEV 空间, 然后将二者在 BEV 空间上进行融合, 这种情况下雷达和视觉没有了主次依赖, 从而能够实现近似后融合的灵活性; 3. 参与了 BEVHeight[13,14] 框架的设计, 传统基于深度估计的方法深度预测对相机参数和外部参数的变化非常敏感, 我们通过回归目标到地面的高度, 并经过相似三角形间接转换为深度来缓解优化过程, 特别是外参的噪声干扰下, 检出稳定性得到了很大的改善; 4. 参与设计了 Opensight[11], 传统检测主要集中在封闭集合中, 我们设计了一个基于 LiDAR 的开放词汇检测框架, 我们先在图像中为通用对象生成 2D 框, 然后将这些 2D 框与 LiDAR 点一起投影回 LiDAR 空间, 估计相应的 3D 框, 同时开发了一个跨模态对齐和融合模块, 将 3D 特征与 2D 图像嵌入对齐, 然后融合对齐的 3D-2D 特征进行语义解码。

### 暗物智能 | 自动机器学习实验室 | 算法实习生

2021.03-2021.08

- **神经网络结构搜索和自监督学习**: 1. 神经网络结构搜索: 提出了一种分块自监督神经网络结构搜索方法 BossNAS[20], 它解决了以往神经网络结构搜索方法中由于权重共享空间过大和监督偏差而导致的模型排序不准确的问题; 2. 自监督学习: 为了避免了自监督学习中繁琐的人工视图设计, 提出了一种自动化学习自监督视图的方法 AutoView[22], 对抗学习最小化互信息的自监督视图, 同时提出了自约束损失, 联合学习最优互信息的自监督视图。

## 论文发表 [谷歌学术: [Link](#) | 引用: 1000+]

### 自动驾驶 | 重建-生成

[1] OmniGen: Unified Multimodal Sensor Generation for Autonomous Driving

Under review

Tao Tang, Enhui Ma, Xia Zhou, Letian Wang, Tianyi Yan, Xueyang Zhang, Kun Zhan, Peng Jia, XianPeng Lang, Jia-Wang Bian, Kaicheng Yu, Xiaodan Liang

[2] RoboPearls: Editable Video Simulation for Robot Manipulation *Under review*  
**Tao Tang\***, Likui Zhang\*, Youpeng Wen, Kaidong Zhang, Jia Wang Bian, Xia Zhou, Tianyi Yan, Kun Zhan, Peng Jia, Hefeng Wu, Liang Lin, Xiaodan Liang

[3] Geometry-Aligned Multimodal Implicit Field for LiDAR-Camera Joint Synthesis, [paper](#), [code](#) *CVPR 2024 Highlight*  
**Tang Tao**, Guangrun Wang, Yixing Lao, Peng Chen, Jie Liu, Liang Lin, Kaicheng Yu, Xiaodan Liang

[4] LiDAR-NeRF: Novel LiDAR View Synthesis via Neural Radiance Fields, [paper](#), [code](#) *ACM MM 2024 Oral*  
**Tang Tao**, Longfei Gao, Guangrun Wang, Peng Chen, Dayang Hao, Xiaodan Liang, Mathieu Salzmann, Kaicheng Yu

[5] LiT: Unifying LiDAR” Languages” with LiDAR Translator, [paper](#), [code](#) *NeurIPS 2024*  
Yixing Lao, **Tang Tao**, Xiaoyang Wu, Peng Chen, Kaicheng Yu, Hengshuang Zhao

[6] Unleashing Generalization of End-to-End Autonomous Driving with Controllable Long Video Generation, [paper](#) *Under review*  
Enhui Ma, Lijun Zhou, **Tao Tang**, Zhan Zhang, Dong Han, Junpeng Jiang, Kun Zhan, Peng Jia, Xianpeng Lang, Haiyang Sun, Di Lin, Kaicheng Yu

[7] LiDAR-GS: Real-time LiDAR Re-Simulation using Gaussian Splatting, [paper](#) *Under review*  
Qifeng Chen, Sheng Yang, Sicong Du, **Tao Tang**, Peng Chen, Yuchi Huo

自动驾驶 | 感知-规控-检索

[8] S2-Track: A Simple yet Strong Approach for End-to-End 3D Multi-Object Tracking, [paper](#) *ICML 2025*  
**Tao Tang\***, Lijun Zhou\*, Pengkun Hao, Zihang He, Kalok Ho, Shuo Gu, Zhihui Hao, Haiyang Sun, Kun Zhan, Peng Jia, Xianpeng Lang, Xiaodan Liang

[9] BEV-TSR: Text-Scene Retrieval in BEV Space for Autonomous Driving, [paper](#) *AAAI 2025*  
**Tang Tao\***, Dafeng Wei\*, Zhengyu Jia\*, Tian Gao\*, Changwei Cai, Chengkai Hou, Peng Jia, Kun Zhan, Haiyang Sun, Jingchen Fan, Yixing Zhao, Fu Liu, Xiaodan Liang, Xianpeng Lang, Yang Wang

[10] Making large language models better planners with reasoning-decision alignment, [paper](#), [code](#) *ECCV 2024 Oral*  
Zhijian Huang\*, **Tang Tao\***, Shaoxiang Chen, Sihao Lin, Zequn Jie, Lin Ma, Guangrun Wang, Xiaodan Liang

[11] Opensight: A simple open-vocabulary framework for lidar-based object detection, [paper](#), [code](#) *ECCV 2024*  
Hu Zhang, Jianhua Xu, **Tang Tao**, Haiyang Sun, Xin Yu, Zi Huang, Kaicheng Yu

[12] Benchmarking the Robustness of LiDAR-Camera Fusion for 3D Object Detection, [paper](#), [code](#) *CVPRW 2023*  
Kaicheng Yu\*, **Tang Tao\***, Hongwei Xie, Zhiwei Lin, Tingting Liang, Bing Wang, Peng Chen, Dayang Hao, Yongtao Wang, Xiaodan Liang

[13] BEVHeight: A Robust Framework for Vision-based Roadside 3D Object Detection, [paper](#), [code](#) *CVPR 2023*  
Lei Yang, Kaicheng Yu, **Tao Tang**, Jun Li, Kun Yuan, Li Wang, Xinyu Zhang, Peng Chen

[14] Bevheight++: Toward robust visual centric 3d object detection, [paper](#), [code](#) *TPAMI*  
Lei Yang, **Tao Tang**, Jun Li, Peng Chen, Kun Yuan, Li Wang, Yi Huang, Xinyu Zhang, Kaicheng Yu

[15] BEVFusion: A Simple and Robust LiDAR-Camera Fusion Framework, [paper](#), [code](#) *NeurIPS 2022*  
Tingting Liang, Hongwei Xie, Kaicheng Yu, Zhongyu Xia, Zhiwei Lin, Yongtao Wang, **Tao Tang**, Bing Wang, Zhi Tang

计算机视觉 | 3D 大模型

[16] UniGS: Unified Language-Image-3D Pretraining with Gaussian Splatting, [paper](#) *ICLR 2025*  
Haoyuan Li, Zhou Yanpeng, **Tang Tao**, Jifei Song, Yihan Zeng, Michael Kampffmeyer, Hang Xu, Xiaodan Liang

[17] Does Your 3D Encoder Really Work? When Pretrain-SFT from 2D VLMs Meets 3D VLMs , [paper](#) *Under review*  
Haoyuan Li, Yanpeng Zhou, Yufei Gao, **Tang Tao**, Jianhua Han, Zhenyu Chen, Jiawang Bian, Hang Xu, Xiaodan Liang

[18] Pruning, Sampling & Merging: Enabling LLMs with Awareness to Point Clouds via 2D Visual Encoders *Under review*  
Yufei Gao, Haoyuan Li, **Tang Tao**, Jiawang Bian, Zequn Jie, Xiaodan Liang

计算机视觉 | 基础表征

[19] MLP Can Be A Good Transformer Learner, [paper](#), [code](#) *CVPR 2024 Best paper candidate*  
Sihao Lin, Pumeng Lyu, Dongrui Liu, **Tang Tao**, Xiaodan Liang, Andy Song, Xiaojun Chang

[20] Exploring Hybrid CNN-transformers with Block-wisely Self-supervised Neural Architecture Search, [paper](#), [code](#) *ICCV 2021*  
Changlin Li, **Tao Tang**, Guangrun Wang, Jiefeng Peng, Bing Wang, Xiaodan Liang, Xiaojun Chang

[21] BossNAS Family: Block-wisely Self-supervised Neural Architecture Search, [paper](#), [code](#) *TPAMI*  
Changlin Li, Sihao Lin, **Tao Tang**, Guangrun Wang, Mingjie Li, Zhihui Li, Xiaojun Chang

[22] AutoView: Learning Self-Regularized Adversarial Views for Self-Supervised Vision Transformers, [paper](#), [code](#) *Under review*  
**Tao Tang**, Changlin Li, Guangrun Wang, Kaicheng Yu, Xiaojun Chang, Xiaodan Liang.

获奖荣誉

研究生一等奖学金	中山大学，2022 – 2025
辽宁省优秀毕业生	辽宁省，2021
国家奖学金，国家励志奖学金	中国教育部，2017 – 2020
玉兰奖学金，一等奖学金，校优秀三好学生，校自立自强标兵	大连理工大学，2017 – 2020
国际大学生数学建模竞赛（MCM/ICM）国际一等奖	COMAP，2019
大连市第二十六届大学生数学竞赛一等奖	大连市，2017