

# 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 | 硕博连读

2021.09–2026.06

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, 具身智能。有多模态, NeRF&3DGS, 生成模型, 大模型的经验。

发表顶会/刊论文 16 篇, 其中一作 7 篇 (2 篇 Oral, 1 篇 Highlight), 谷歌学术引用 1100+。

### 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

- 自动驾驶场景多模态重建: 1) 提出了首个基于神经渲染的 LiDAR 仿真生成器 LiDAR-NeRF[4], 可端到端合成新视角的点云数据。 2) 推出多模态统一渲染模型 AlignMiF[3], 解决 LiDAR 与图像间的不一致问题, 实现可微分的多模态神经场建模。 3) 结合 3DGS, 提出 LiDAR-GS[7], 显著提升渲染速度与质量。 4) 推出 RoboPearls[2], 基于 GS 构建语义增强的动态场景重建, 并结合大语言模型支持自然语言编辑, 有效增强下游机器人操纵性能。
- 自动驾驶场景多模态生成: 1) 参与设计 Delphi[6], 一种基于扩散模型的长视频生成方法, 引入跨视角共享噪声机制与特征对齐模块以增强空间与时间一致性。 2) 提出首个端到端多模态场景生成框架 OmniGen[1], 在统一 BEV 表征中融合 LiDAR 与图像, 通过扩散模型生成多模态数据。 3) 提出 BEV-TSR[9], 首个基于 BEV 空间的文本场景检索框架, 结合知识图谱与大语言模型增强语义理解, 实现高精度多模态检索。
- 端到端大模型: 1) 提出基于视觉语言模型 (VLM) 的端到端规划方法 RDA-Drive[10], 针对现有方法忽视因果推理与决策间不一致的问题, 引入排序对齐损失 (Rank-Alignment Loss), 强化 CoT 与决策输出的一致性, 实现领先的自动驾驶规划性能。 2) 构建端到端 3D 多目标跟踪模型 S2-Track[8] (持续领先 nuScenes 排行榜): Query 初始化: 结合 2D 目标位置与深度信息估计初始 3D 框; Query 优化: 引入不确定性感知解码器, 将注意力建模为高斯分布以量化跟踪不确定性; Query 训练: 通过向真实目标框添加噪声并进行去噪训练, 提升匹配鲁棒性与稳定性。

### 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: 研发鲁棒可靠的多模态感知算法: 1) 主导构建 RobustBenchmark[12], 收集 BEV 感知中常见噪声干扰样例, 开发系统化鲁棒性测试工具包; 在 nuScenes 和 Waymo 上建立新鲁棒性基准, 全面评估主流 BEV 融合方法的抗干扰性能。 2) 参与 BEVFusion[15] 框架设计: 采用独立网络分别编码雷达与图像特征, 统一投影至 BEV 空间再进行融合, 实现近似后融合的结构, 打破视觉对雷达的主导依赖。 3) 参与设计 BEVHeight[13,14]: 为提升检测在参数扰动下的稳定性, 提出以目标高度回归替代直接深度预测, 通过几何转换间接获取深度, 显著缓解相机参数扰动对性能的影响。 4) 参与开发 Opensight[11]: 提出基于 LiDAR 的开放词汇检测框架。通过图像生成通用 2D 检测框并回投至 LiDAR 空间估计 3D 位置; 设计跨模态对齐与融合模块, 将 3D 与 2D 特征对齐后进行语义解码, 实现开放类目标检测能力。

### 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) 提出结构搜索方法 BossNAS[20], 通过分块自监督机制缓解权重共享空间过大和监督偏差问题, 显著提升模型排序准确性与搜索效果。 2) 设计自动化自监督视图生成方法 AutoView[22], 避免繁琐手工视图设计, 以对抗方式最小化视图间互信息; 提出自约束损失函数, 联合优化互信息以提升自监督学习。

## 论文发表 [谷歌学术: [Link](#) | 引用: 1100+]

### 自动驾驶 | 重建-生成

- [1] OmniGen: Unified Multimodal Sensor Generation for Autonomous Driving, [paper](#) *ACM MM 2025*  
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, [paper](#) *ICCV 2025*  
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](#)  
Yixing Lao, **Tang Tao**, Xiaoyang Wu, Peng Chen, Kaicheng Yu, Hengshuang Zhao

*NeurIPS 2024*

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

*Under review*

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

*Under review*

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](#)  
**Tao Tang\***, Lijun Zhou\*, Pengkun Hao, Zihang He, Kalok Ho, Shuo Gu, Zhihui Hao, Haiyang Sun, Kun Zhan, Peng Jia, Xianpeng Lang, Xiaodan Liang

*ICML 2025*

[9] BEV-TSR: Text-Scene Retrieval in BEV Space for Autonomous Driving, [paper](#)  
**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

*AAAI 2025*

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

*ECCV 2024 Oral*

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

*ECCV 2024*

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

*CVPRW 2023*

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

*CVPR 2023*

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

*TPAMI*

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

*NeurIPS 2022*

计算机视觉 | 3D 大模型

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

*ICLR 2025*

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

*Under review*

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

*Under review*

计算机视觉 | 基础表征

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

*CVPR 2024 Best paper candidate*

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

*ICCV 2021*

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

*TPAMI*

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

*Under review*

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

技能

编程语言与工具：Python, PyTorch, Git