Qin You

Last updated in May 2026

PH.D. CANDIDATE · COMPUTER SCIENCE · NATIONAL UNIVERSITY OF SINGAPORE

e0962995@u.nus.edu | +65 8039 5597 | Homepage | Google Scholar

EDUCATION

National University of Singapore — Aug 2024 – Present
Ph.D. in Computer Science · *Advisor: Prof. Roger Zimmermann*

National University of Singapore — Aug 2022 – Jan 2024
Master of Computing, General Track — *GPA **4.50 / 5.0***

University of Electronic Science and Technology of China — Sep 2018 – Jun 2022
B.Sc. in Mathematics for Information and Computing Science — *GPA **3.71 / 4.0***

RESEARCH EXPERIENCE

Tencent Hunyuan, Application Center — Nov 2025 – Present
Research Intern · *Advisors: Chunyu Wang & Linqing Wang* — *Shenzhen, China*

(1) SOAR: Self-Correction for Optimal Alignment in Rectified-Flow Diffusion
Project Page · Report

- Proposed a scalable, *reward-free* post-training framework that targets exposure bias in the denoising trajectory by sampling on-trajectory states, performing a stop-gradient CFG rollout with the current model, and supervising the denoiser with analytical correction targets.
- Designed an on-policy, dense, per-timestep objective that subsumes the standard SFT loss and remains compatible with subsequent RL alignment.
- On SD3.5-Medium: GenEval **0.70** → **0.78** and OCR **0.64** → **0.67** over SFT, with simultaneous gains on every DrawBench preference and quality metric; surpasses Flow-GRPO on aesthetic and CLIPScore tasks despite using *no* reward model.
- Open-sourced as HY-SOAR (400+ GitHub stars).

(2) JAVRL: Joint Audio-Video RL for Multi-Speaker Lip-Sync Alignment

- Contributed to project design; specifically designed a **contrastive speaker-sync reward** for dual/multi-speaker scenes that compares SyncNet scores across visible faces to enforce correct who-speaks-what assignment, requiring no annotations or paired lip-sync data.

Intelligent Machine Perception Lab, Singapore University of Technology and Design — Mar 2024 – Aug 2024
Research Associate · *Advisor: Prof. Na Zhao* — *Singapore*

(1) Fully-Sparse Multi-modal 3D Object Detection with Dynamic Prompting

- Designed a fully-sparse architecture fusing LiDAR point clouds with RGB images to improve 3D detection efficiency and accuracy.
- Implemented a dynamic-prompt mechanism that adaptively guides the detector's attention across diverse scenes.

(2) Pre-trained Diffusion for Single-view 3D Scene Generation

- Investigated pre-trained diffusion priors for generating 3D Gaussian-Splatting representations from a single image.
- Prototyped an image-to-3D pipeline that hallucinates occluded content via diffusion and converts it into a continuous representation for novel-view synthesis.

Next++ Sea Joint Lab, National University of Singapore — Sep 2022 – Mar 2024
Research Intern · *Advisor: Prof. Wei Ji* — *Singapore*

(1) Multi-modal Information Retrieval for Panoptic Scene Graph Generation (PSG)

- Developed representation-approximation and invariant-learning techniques that find "invariant" predicate representations across visual semantics, bridging multimodal data for the PSG task.
- Introduced a coreset-based formulation to clean multimodal datasets; recall improved by ∼**50%** over prior methods on semantic-relation recognition.
- The proposed framework generalises naturally to arbitrary modality combinations.

(2) Multi-modal Understanding for Video Moment Retrieval (VMR)

- Designed **MRTNet**, a plug-and-play module enforcing feature consistency across multi-resolutions and queries; significant gains at marginal compute cost.

- Built **IntroDis**, a dual-teacher framework integrating in-distribution and OOD perspectives with counterfactual reasoning for robust OOD generalisation.
- Reconstructed a new benchmark and proposed a complete framework for **Described Spatial-Temporal Video Detection (DSTVD)**, grounding arbitrary object tubelets from natural-language descriptions.

SURVEY

» **Audio-Visual Intelligence in Large Foundation Models** ★11

You Qin, Kai Liu, Shengqiong Wu, Kai Wang, Shijian Deng, Yapeng Tian, Junbin Xiao, Yazhou Xing, Yinghao Ma, Bobo Li, Roger Zimmermann, Lei Cui, Furu Wei, Jiebo Luo, Hao Fei

arXiv:2605.04045 · 56 pages, 16 figures · Submitted 5 May 2026 [**Survey**]

FULL PUBLICATIONS

** indicates equal contribution. Underlined name is the candidate.*

» **SOAR: Self-Correction for Optimal Alignment and Refinement in Diffusion Models** ★404 🤗

You Qin, Linqing Wang, Hao Fei, Roger Zimmermann, Liefeng Bo, Qinglin Lu, Chunyu Wang

Conference on Neural Information Processing Systems (NeurIPS) 2026 [**Under Review**]

» **Geometry over Density: Few-Shot Cross-Domain OOD Detection**

Shawn Li*, **You Qin***, Jiate Li, Charith Peris, Lisa Bauer, Roger Zimmermann, Yue Zhao

Conference on Neural Information Processing Systems (NeurIPS) 2026 [**Under Review**]

» **Contextual Hashing Meets Lightweight Convolution: Accelerating Retrieval and Refining Localization for Video Corpus Moment Retrieval**

Mingjin Kuai, Jin Peng, Zheqi Lv, **You Qin**, Zhan Yang, Zhen Zhang, Wei Zhou, Wei Ji

IEEE Transactions on Image Processing (TIP) [**Under Review**]

» **SRDiff: A Cross-Modal Diffusion Model for Satellite-to-Radar Translation in Precipitation Nowcasting**

You Qin, Jinming Cao, Ting Wang, Yifang Yin, Li Li, Shili Xiang, Ying Zhang, Roger Zimmermann

IEEE Transactions on Geoscience and Remote Sensing (TGRS) 2026

» **Dynamic Graph-enhanced Event Refinement for Temporal Sentence Grounding of Micro-moments**

Mingjin Kuai*, **You Qin***, Xiang Fang, Yiming Wu, Wei Ji, Roger Zimmermann

IEEE Transactions on Multimedia (TMM) 2026

» **Grounding is All You Need? Dual Temporal Grounding for Video Dialog**

You Qin, Wei Ji, Xinze Lan, Hao Fei, Xun Yang, Dan Guo, Roger Zimmermann, Lizi Liao

IEEE Transactions on Multimedia (TMM) 2026

» **InstaFlow-Pan: One-Step Flow Matching for High-Fidelity Pansharpening**

Qian Liu, **You Qin**, Zhiyuan Li, Xiangyong Cao, Junmin Liu

IEEE Transactions on Geoscience and Remote Sensing (TGRS) 2026

» **FSGformer: Frequency Separation and Guidance Transformer for Pansharpening**

Qian Liu, Xiangyu Zhao, **You Qin**, Lanyu Li, Junmin Liu

IEEE Transactions on Geoscience and Remote Sensing (TGRS) 2025

» **Secure On-Device Video OOD Detection Without Backpropagation** ★14

Li Li, Peilin Cai, Yuxiao Zhou, Zhiyu Ni, Renjie Liang, **You Qin**, Yi Nian, Zhengzhong Tu, Xiyang Hu, Yue Zhao

International Conference on Computer Vision (ICCV) 2025

» **Generalized Video Moment Retrieval**

You Qin, Qilong Wu, Yicong Li, Wei Ji, Li Li, Pengcheng Cai, Lina Wei, Roger Zimmermann

International Conference on Learning Representations (ICLR) 2025

» **Described Spatial-Temporal Video Detection**

Wei Ji, Xiangyan Liu, Yingfei Sun, Jiajun Deng, **You Qin**, Ammar Nuwanna, Mengyao Qiu, Lina Wei, Roger Zimmermann

arXiv preprint arXiv:2407.05610 [**Preprint**]

» **Unveiling Causalities in SAR ATR: A Causal Interventional Approach for Limited Data**

Chenwei Wang, Xin Chen, **You Qin**, Siyi Luo, Yulin Huang, Jifang Pei, Jianyu Yang

IEEE Geoscience and Remote Sensing Letters (GRSL) 2024

» **Panoptic Scene Graph Generation with Semantics-prototype Learning** ★35

Li Li, Wei Ji, Yiming Wu, Mengze Li, **You Qin**, Lina Wei, Roger Zimmermann

AAAI Conference on Artificial Intelligence (AAAI) 2024

» **MRTNet: Multi-Resolution Temporal Network for Video Sentence Grounding**

Wei Ji, **You Qin**, Long Chen, Yinwei Wei, Yiming Wu, Fangfang Wang, Roger Zimmermann

IEEE International Conference on Acoustics, Speech & Signal Processing (ICASSP) 2024

» **Domain-wise Invariant Learning for Panoptic Scene Graph Generation**
Li Li, **You Qin**, Wei Ji, Yuxiao Zhou, Roger Zimmermann
IEEE International Conference on Acoustics, Speech & Signal Processing (ICASSP) 2024

» **Biased-Predicate Annotation Identification via Unbiased Visual Predicate Representation**
Li Li*, Chenwei Wang*, **You Qin**, Wei Ji, Renjie Liang
ACM International Conference on Multimedia (ACM MM) 2023

» **Causal SAR ATR with Limited Data via Dual Invariance**
Chenwei Wang, **You Qin**, Li Li, Siyi Luo, Yulin Huang, Jifang Pei, Ying Zhang, Jianyu Yang
arXiv preprint arXiv:2308.09412 **[Preprint]**

COURSEWORK & AWARDS

SELECTED MODULES

Intelligent Robots: Algorithms & Systems · AI Planning & Decision Making · Uncertainty Modelling in AI · Theory & Algorithms for Machine Learning.

HONOURS & AWARDS

- Honourable Mention, *Interdisciplinary Contest in Modeling* (ICM).
- 2nd Prize, *National College Student Mathematical Contest in Modeling* (China).
- 2nd Prize, *UESTC 18th Programming Contest.*
- 3rd Prize, *11th National College Students Mathematics Competition* (China).

ACADEMIC SERVICE

CONFERENCE REVIEWER

- **ACM Multimedia (ACM MM)** — 2023, 2024 **[Outstanding Reviewer, 2024]**
- **International Conference on Computer Vision (ICCV)** — 2025
- **International Conference on Learning Representations (ICLR)** — 2025, 2026
- **Conference on Neural Information Processing Systems (NeurIPS)** — 2025, 2026
- **IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)** — 2026
- **European Conference on Computer Vision (ECCV)** — 2026

WORKSHOP ORGANIZER

- **JAV-CG: 1st International Workshop on Joint Audio-Video Comprehension and Generation** — co-located with ACM Multimedia 2026, Rio de Janeiro. · Homepage · Code

TEACHING

Teaching Assistant *National University of Singapore*

- CS4243 – Computer Vision and Pattern Recognition *Aug – Nov 2025*
- CS5346 – Information Visualisation *Jan – May 2025*

TECHNICAL SKILLS

LANGUAGES	C / C++ · Python · Java · JavaScript / TypeScript · HTML / CSS · LaTeX
AI / ML	PyTorch · PyTorch Lightning · TensorFlow · NumPy · Pandas
WEB	Vue.js · React · Node.js
TOOLS	Git · Unix Shell · MySQL · Docker