Junwen Chen

585-360-6687 • junwenchen.swjtu@gmail.com • junwenchen.github.io

Work Experiences

Amazon AGI - Foundation **Mar 2025 – present**

Applied Scientist @ Nova Vision Skills Team

Project: Video/Omni understanding data for Amazon Nova2 Lite/Pro/Omni

- Develop new feature temporal/timestamp understanding.
  - Built training data pipeline: Diversified video sourcing, scalable training data generation across PT/SFT/RL, and balanced data recipe design, with PySpark.
  - Built timestamp video token sequence construction module in training framework.
  - Addressed the temporal localization across multiple customer domains and achieved state-of-the-art on temporal grounding QVHighlight/Charades benchmarks.
- Support new feature video+audio cross-modal understanding.
  - Curated a novel PT/SFT data generation pipeline to address video+audio (omni) retrieval and video chaptering and time-aware video caption, with PySpark.
  - Addressed the new capability and improve general video understanding ability i.e. VideoQA video-mme benchmark.
- Solve a variety of new tasks for Nova 2 to support customer interested domains, including: Cinematic video understanding (audio description, visual effect detection, knowledge-enhanced recognition e.g. character recognition), Sports video understanding (counting, highlight retrieval), Surveillance videos (anomaly activity), etc.

Project: Text retrieval for Amazon Nova Multimodal Embedding

- Build a data generation pipeline to improve expert domains and multilingual and design a new data recipe to improve text-to-text retrieval performance while maintaining the other modality (image, video, audio, document) unimodal, cross-modal performance.

Amazon Web Services **Apr 2024 – Mar 2025**

Applied Scientist @ AWS AI Lab

Project: Multi-modal insight extraction system for Amazon Bedrock Data Automation

- Design and build scalable data generation pipeline for multiple multi-modal tasks, i.e. image/video summary, blueprint key information extraction (KIE), taxonomy-based tagging, based on a series of LLMs and in-house vision experts.
- Build post-training framework based on Qwen2.5VL.
- Build in-house evaluation using LLM ratings.

Research Interests

General interests LLM, Multi-modal Understanding, Data-centric AI.
Specific interests Omni Models (Video/Audio), Computer Use, Tool Calling, Video Editing, Streaming Understanding, Trustworthy Computer Vision.

Skills

PyTorch, PySpark, Pandas, Docker, Ray, Git, Kubernetes cluster, AWS services (S3, EC2, Batch, EMR, ECR), Azure ML

Education

- **Michigan State University** — **Aug 2022 – Apr 2024**
  Ph.D in Computer Science — *Advisor: Yu Kong*
  Thesis: Action Modeling in Long-form Videos
- **Rochester Institute of Technology** — **Sep 2018 – Jul 2022**
  Ph.D student in Computer Science — *Advisor: Yu Kong*
- **Southwest Jiaotong University** — **Sep 2011 – Jun 2018**
  B.S & M.S. in Electrical Engineering — *Advisor: Zhigang Liu*

Publications

[0] Two papers in submission to 2026 computer vision and NLP venues.

[1] **Junwen Chen**, Jie Zhu, Yu Kong. Action Temporality Modeling for Video Question Answering, ACM Multimedia 2023.

[2] **Junwen Chen***, Gaurav Mittal*, Ye Yu, Yu Kong†, Mei Chen†. GateHUB: Gated History Unit with Background Suppression for Online Action Detection. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2022.

[3] **Junwen Chen**, Yu Kong. Explainable Video Entailment with Grounded Visual Evidence. *The International Conference on Computer Vision (ICCV)*, October 2021.

[4] **Junwen Chen**, Wentao Bao, Yu Kong. Group Activity Prediction with Sequential Relational Anticipation Model. *The European Conference on Computer Vision (ECCV)*, August 2020.

[5] **Junwen Chen**, Wentao Bao, Yu Kong. Activity-driven Weakly-Supervised Spatio-Temporal Grounding from Untrimmed Videos. *ACM Multimedia*, October 2020.

[6] **Junwen Chen**, Haiting Hao, Hanbin Hong, Yu Kong. RIT-18: A Novel Dataset for Compositional Group Activity Understanding. *CVPR Women in Computing Workshop*, June 2020.

[7] **Junwen Chen**, Zhigang Liu, Hongrui Wang, Alfredo Nunez, Zhiwei Han. Automatic Defect Detection of Fasteners on the Catenary Support Device Using Deep Convolutional Neural Network. *IEEE Transactions on Instrumentation and Measurement*, vol. 67, no. 2, Feb. 2018.

[8] **Junwen Chen**, Zhigang Liu, Hongrui Wang, Kai Liu. High-speed railway catenary components detection using the cascaded convolutional neural networks. *IEEE International Conference on Imaging Systems and Techniques*, October 2017.

[9] Hongrui Wang, Alfredo Núñez, Rolf Dollevoet, Zhigang Liu, **Junwen Chen**. Intelligent condition monitoring of railway catenary systems: A Bayesian network approach. In Proc. 25th Int. Symp. Dyn. Vehicles Roads Tracks (IAVSD), Dec. 2017.
* indicates equal contribution. † indicates equal advising.

Research Projects at MSU&RIT

Learning action representations for VideoQA
- Proposed a novel empirical study on video representations for motion-plentiful videos.
- Proposed a novel contrastive learning and a loss function to learn faithful action representations.
- Paper accepted to MM2023.

Activity Anticipation of Multiple People

- Proposed a sequential relational anticipation framework to predict the activity of a conceptual group of people when only the beginning frames of the activity are observed.
- Marginally outperformed the best existing action prediction methods and also explicitly predicted the trajectories of multiple people in the unseen frames of the activity as a by-product.
- Established a new dataset for compositional group activity recognition and long-duration anticipation.
- Paper accepted to ECCV 2020.

Weakly Supervised Video Object Grounding

- Proposed a video object grounding framework using the off-the-shelf video captions but without extensively finely-annotated object tracklets, which greatly decreases the labeling cost for real-world applications.
- Built a spatio-temporal context encoding model to capture the temporal patterns of object interactions, a good way to ground objects driven by activity information in captions, which boosted state-of-the-art methods by 2%.
- Paper accepted to ACM Multimedia 2020.

Explainable Video Entailment

- Proposed an explainable model to improve video entailment by grounding the textual description to visual content and regularizing the training by highlighting the contradition textual content.
- Paper accepted to ICCV 2021.

Internship Experiences

AWS AI Lab **May 2023 – Aug 2023**
Applied Science Intern *Mentor: Onkar Dabeer*

Large Language Model for VideoQA

- Proposed a slow-fast based framework to bridge image/video foundational models with large language models for long-form videoQA.
- Achieved state-of-the-art on widely-used VideoQAs, including NextQA, STAR,

Microsoft Cloud & AI **May 2021 – Aug 2021**
Research Intern *Mentor: Mei Chen*

Streaming and Untrimmed Video Understanding

- Built a novel transformer-based framework to detect the ongoing action in a video stream. Developed a learning objective to better separate the actions and context.
- Paper accepted to CVPR 2022.

Selected Awards

CVPR Doctoral Consortium Award	*Jun. 2023*
Conference DEI travel grant to attend CVPR 2023, Vancouver, BC	*Jun. 2023*
Conference travel grant to attend CVPR 2022, New Orleans, LA	*Jun. 2022*
CVPR Outstanding Reviewer Recognition	*Jun. 2021*
Conference travel award to attend AAAI 2020, New York, NY	*Feb. 2020*
CVPR Women in Computing Workshop 2020 Student Award	*Jun. 2020*

Academic Service

Conference Reviewer: CVPR 2021-, ECCV 2022, ICCV 2021, ACM Multimedia 2019-2022

Journal Reviewer: IEEE Trans. on Instrum. and Meas., IEEE Trans. on Intelligent Transportation System, Applied Soft Computing, Measurement
Student Volunteer: AAAI 2020, CCC 2016