Junjie Jiang（江俊杰）

Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences
Email: jj.jiang@siat.ac.cn & dringel.chiang@gmail.com | Homepage: https://junjiechiang.github.io

ABOUT

I am currently a research assistant at Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, supervised by Prof. Zhile Yang and A/Prof. Chengke Wu. Previously, I graduated from Shenyang Ligong University with a bachelor's degree in 2022. My current research is the integration of large pre-trained language models and knowledge engineering and its application in knowledge-intensive domains.

EDUCATION

Shenyang Ligong University　　　　　　　　　　**B. Eng**　　　　　　　　　**Sep.2018-Jun.2022**

- Core Curriculums: Advanced Mathematics, Linear Algebra, Probability Theory, C language programming, Digital Image Processing, Robot Dynamics, ROS Robot System
- Thesis: Research and application of face recognition based on deep learning

Shenzhen Institute of Advanced Technology, CAS　　**Visiting Student**　　　　　**Jun.2022-Present**

- Research direction: the intersection of natural language processing, knowledge engineering, computer vision and machine learning, and their applications in areas such as smart construction and smart grids.

PUBLICATIONS

- **Jiang, J.**, Yang, Z., Wu, C., Guo, Y., Yang, M., & Feng, W*. (2023). A compatible detector based on improved YOLOv5 for hydropower device detection in AR inspection system. *Expert Systems with Applications*, 225, 120065. (**JCR Q1 IF 8.5**)

- Yang, M., Wu, C., Guo, Y., He, Y., Jiang, R., **Jiang, J.**, Yang, Z*. (2024). A teacher-student deep learning strategy for extreme low resolution unsafe action recognition in construction projects. *Advanced Engineering Informatics*, 59, 102294. (**JCR Q1 IF 8.8**)

- **Jiang, J.**, Wu, C.*, Liu, Z., Guo, Y., Sun, W., & Yang, Z. (2024). Ontology-based distant supervision for extracting entity-property relations in construction documents. *Automation in Construction*. (**JCR Q1 IF 10.3 Under Review AUTCON-S-24-00151**)

- Wu, C., **Jiang, J.**, Wu, X., Li, X., Guo, Y., & Yang, Z*. (2024). Large language model driven heterogenous information integration, searching, and interaction in construction management. *Computers in Industry*. (**JCR Q1 IF 10 Pending for Submission**)

SELECTED PATENTS

- Yang, Z., **Jiang, J.**, Guo, Y., Liu, X., Wu, C. Machine Vision-based Battery Surface Defect Detection Method and System and Related Device. WIPO (PCT) Patent, **WO2023226357A1**, publication.

- Yang, Z., **Jiang, J.**, Guo, Y., Liu, X., Wu, C. Machine Vision-based Battery Surface Defect Detection Method and System and Related Device. National Invention Patent, **CN114972258B**, authorization.

- Yang, Z., **Jiang, J.**, Liu, X., Guo, Y., Wu, C. A Real-time Optimization Control Method for Charging and Discharging States of a Hybrid Energy Storage System. National Invention Patent, **CN115313447A**, authorization.

- Guo, Y., **Jiang, J.**, Wu, C., Yang, Z., Hu, T. Method, System, and Related Equipment for Defect Detection Based on Battery Surface Images. National Invention Patent, **CN115272330A**, authorization.

- Guo, Y., Zhu, J., **Jiang, J.**, Hu, T., Wang, L., Lu, J., He, Y., Wei, G. A Method and System for Predicting and Compensating the Spindle Error of CNC Machine Tools. National Invention Patent, **CN114690706B**, authorization.

RESEARCH EXPERIENCE

Research Assistant   **Jun.2023-Present**

Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences   *Shenzhen*

Domain knowledge enhanced large language model for the construction industry (Work in Progress)

- Inspired by HuatuoGPT, FinGPT, and ChatLaw, etc., the research mainly involves 1) forming chat dataset by extracting Q&A from exam papers with cutting-edge OCR tools, 2) enriching the answers of exam papers through knowledge distillation with large language model (LLM), 3) pre-training base models with fine-tuning strategies, e.g., LoRa, SFT, etc., and 4) filling the gap of a few LLMs in the construction industry.

- This research is being conducted under the joint collaboration of the Institute of Computing Technology, CAS, and Shandong University.

Knowledge-enhanced relation extraction using large language model as distant supervisor (Work in Progress)

- The research is based on the previously completed study, which focuses on utilizing LLM with the most advanced knowledge integration strategies, e.g., knowledge graph, retrieval augmented generation, etc., to improve relation annotation during distant supervision.

- The aims are to 1) explore the possibility of knowledge enhanced LLM as domain data annotator, 2) improve the accuracy of relation labelling during the semi-supervised annotation process, and 3) alleviate the extensive human intervention in domain relation extraction dataset development.

Large language model driven heterogenous information integration, searching, and interaction in construction management (Funded by Young Scientists Fund of the National Natural Science Foundation of China)

- The research aims to utilize LLM to handle user-system interactions through information hidden in heterogenous sources, e.g., project documents and BIM models.

- Hence, the LLM for Construction Management (LLM4CM) paradigm is proposed to enhance the user-system interaction in CM, which includes: 1) Developed a multi-source vectorized database (i.e., containing information in documents and software), 2) Designed a voting mechanism to search information by evaluating the relevance between the user query and information source and then grouping, and 3) the theoretical LLM usage framework is developed to map LLMs with different scales to the main tasks of CM.

Ontology-based distant supervision for extracting entity-property relations in construction documents (Funded by China Postdoctoral Science Foundation Project)

- Responsible for developing entity-property relation extraction for construction documents using a novel knowledge-based distant supervision framework.

- The aims are to 1) reduce the labelling noise in distant supervision through knowledge integration and semantic similarity, 2) alleviate the cost of developing a large-scale KB covering abundant entities with similar semantics, and 3) avoid the massive manual annotation of supervised learning for model training.

- Hence, the Ontology for Relation Extraction (Ont4RE) approach is proposed, which includes 1) Established an ontology as an external KB, 2) Developed an ontology-based distant supervision strategy for automatic annotation, and 3) Developed BERT-based models pre-trained on self-built construction corpus.

Visiting Student   **May.2022-Jan.2023**

Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences   *Shenzhen*

A compatible detector based on improved YOLOv5 for hydropower device detection in AR inspection system (Funded by University-Enterprise Cooperation Project)

- Responsible for the development of a hydropower device (object) detection algorithm in AR inspection system.

- The aims are to 1) alleviate the computation and usage of the object detection model for deploying it to AR systems, 2) solve the information loss when scale-different feature maps are fused as the person wearing the AR device keeps moving, and 3) optimize hyperparameters for DL model to adapt to working scenarios.

- Hence, the YOLO-Master is proposed, which includes 1) Replaced the original backbone with MobileNetv3 to

reduce the computation and usage, 2) Introduced Coordinate Attention mechanism to enhance the image features, and 3) Optimized hyperparameters configuration with genetic algorithm.

Undergraduate   **Nov.2021-May.2022**

Shenyang Ligong University   *Shenyang*

Research and application of face recognition based on deep learning. (Graduation Thesis)

- The thesis is focused on face recognition using deep learning algorithm.
- Two DL models (i.e., Retinaface and FaceNet) are utilized in the thesis because face recognition involves two phases (i.e., face detection and face recognition).
- The process includes 1) Detected five key points of the face for facial localization, 2) Mapped facial information into feature embeddings in Euclidean space, and 3) Compared embeddings in the input image with the face database through Euclidean distance.

WORK EXPERIENCE

Intern   **May.2022-Jun.2022**

Zhongke Hangmai CNC Software Shenzhen Co., Ltd.   *Shenzhen*

- Participated in a collaborative project with China Southern Power Grid, utilizing AR technology to optimize the operation and maintenance processes in power plants. Responsible for collecting and organizing an image dataset of hydropower devices, supporting object detection and localization in AR technology. Led the development and training of the YOLO-Master detector model, compatible with AR devices. Worked closely with backend engineers to integrate the model's output bounding box information with the backend system.

Database Engineer (Internship)   **May.2020-Aug.2020**

Guangdong Creawor Technology Development Co., Ltd.   *Guiyang*

- Participated in the development of the Data Business Analysis and Marketing Service Support System for China Mobile (Guiyang). Familiar with mainstream relational databases in the industry, and proficient in using SQL and Navicat to provide data support for Guiyang Mobile. Assisted the project team in maintaining the system and the precision marketing platform, meeting the needs of Guiyang Mobile's precision marketing.

AWARD

- SIAT President's Scholarship – Excellent Award (top 5%)   **Jan.2024**
- Director's Innovation Award - Outstanding Graduate Student   **Jan.2024**
- The SEIC Outstanding Contribution Award   **Dec.2022**
- Excellent Undergraduate Thesis of Shenyang Ligong University   **Jun.2022**
- Third Prize in Robot Competition of the School of Mechanical Engineering   **Nov.2019**

ACTIVITIES

- The 25th China Hi-Tech Fair   **Shenzhen, Nov.2023**
- 2023 International Digital Energy Expo.   **Shenzhen, Jul.2023**
- Forum on from GPT-3 to ChatGPT: Innovation and Practice in NLP Applications   **Shanghai, Feb.2023**
- The 24th China Hi-Tech Fair   **Shenzhen, Nov.2022**
- The 3rd International Symposium on New Energy and Electrical Technology   **Henan, Aug.2022**

ACADEMIC SERVICES

Chinese Association of Automation Student Member   IEEE Student Member   Reviewer OICE 2024

ADDITIONAL SKILLS

English Proficiency: CET 6 (473) / IELTS 6(5.5) – *In Preparation*

Technical Skills: Origin, Visio, Python, Latex, Git, Linux, Photoshop, Microsoft 365