


Yue Zhang

✉ skywalkerzhang19@gmail.com  [linkedin.com/YueZhang](https://www.linkedin.com/YueZhang)

Education

The University of Texas at Dallas

Ph.D. in Computer Science (GPA: 4.00 / 4.00). Research Interest: Multi-Modality and NLP

Started on 08/2023

Richardson, Texas, US

The University of Manchester

MSc in Advanced Computer Science (GPA: 3.83 / 4.00, Distinction)

09/2020 - 09/2021

Manchester, England, UK

Zhejiang Normal University

B.E. in Computer Science and Technology (GPA: 3.86 / 4.00, Rank: 1/34)

09/2016 - 06/2020

Jinhua, Zhejiang, CHN

Experience

Multimodal Document Understanding, Hong Kong University of Science and Technology

04/2022 - 03/2023

Research Assistant, Supervisor: Sung Kim, Lucy Park

Guangzhou, Guangdong, CHN

- Conducted research on Information Extraction in Multimodal documentation, training and fine-tuning a series Transformer-based model.
- Introduced a Post Correction Model to improve information extraction results in the field of multimodal document information extraction, improving the accuracy from 68.4438 to 71.6858.
- Proposed Extract Answer Merge Answer (EAMA) in the field of multimodal table information extraction, achieving **third place winner** in the VQAonBD task of the ICDAR competition.

ECG (Electrocardiograph) QRS Feature Disease Recognition

01/2021 - 09/2021

Graduation Project, Supervisor: David Wong

Manchester, England, UK

- Extracted QRS features using three methods of Pan, Lourenco, and Kalidas from PhysioNet, and completed two comparative models.
- Entered features into Adaptive Lead Weighted ResNet with F1, accuracy, recall, precision, increasing the recall and accuracy by 1% for most diseases and by roughly 7% for other diseases.

Twitter Analysis for Tokyo Olympics: Sentiment and Entity Recognition

03/2021 - 04/2021

Course Project

Manchester, England, UK

- Extracted 3400 Olympic-related tweets using TwitterAPI for analysis, performed data cleaning, and visualized attitudes using word clouds.
- Developed the sentiment analysis, Named Entity Recognition, and topic modeling models.

Stereo Vision and 3D Reconstruction

03/2021 - 05/2021

Course Project

Manchester, England, UK

- Implemented edge detection on grayscale stereo images using OpenCV, enhancing feature matching accuracy for subsequent analyses between the left and right views.
- Developed and normalized a disparity map by measuring pixel discrepancies across stereo images, enabling depth estimation for 3D modeling.
- Reconstructed a detailed 3D model of an umbrella from stereo images by leveraging the calculated disparity map combined with precise camera calibration parameters (focal length and baseline), enabling accurate depth perception and virtual reality applications.

Research Papers

Defeasible Visual Entailment for Large Vision-Language Models

Yue Zhang, Liqiang Jing, Vibhav Gogate

AAAI2025

- Proposed novel defeasible visual entailment tasks, including defeasible visual entailment classification and defeasible visual entailment generation, created Defeasible Visual Entailment Benchmark. The current Large Vision-Language Models show a poor performance on the novel defeasible visual entailment task.
- Built a reference-free evaluator based on contrastive learning and multi-task learning. The evaluation performance surpassed the state-of-the-art evaluation metrics.
- Proposed a new reward-driven update optimization method and demonstrate experimentally that our method significantly enhances the quality of generated updates, outperforming state-of-the-art models.
- Responsible for ideas, data collection and annotation, experiment designs, and initial paper writing.

Can Video Large Multimodal Models Think Like Doubters— or Double-Down: A Study on Defeasible Video Entailment

Yue Zhang, JiLei Sun, Yunhui Guo, Vibhav Gogate

Under Review by ICCV

- * Proposed novel tasks to evaluate the defeasibility reasoning capabilities of Video-Large Multimodal Models (VLMMs).
- * Developed a novel Chain of Counterfactual Thought method that integrates visual to improve classification ability of VLMMs and LLM-Guided ASR-Integrated method to improve generation ability of VLMMs.
- * Conducted experiments on state-of-the-art VLMMs and found method could improve the reasoning ability of all the current VLMMs.

Fine-grained and Explainable Factuality Evaluation for Multimodal Summarization

Yue Zhang, Jingxuan Zuo, Liqiang Jing

DoCUI@AAAI2025

- Proposed fine-grained and explainable factuality evaluation frameworks for multimodal summarization under reference-based and reference-free scenarios. The framework can be adapted to Large Vision-Language Models easily.
- Conducted human evaluation regarding faithfulness, and the experimental results show our metrics achieved the best performance compared with the existing generation metrics.
- Responsible for idea discussion, data collection and annotation, experiment design, and initial paper writing.

Machine learning classification of multi-lead ECGs using clinically-relevant features

Yue Zhang, David Wong

Msc Thesis. University of Manchester

- Responsible for idea, experiment design, and paper writing.

Speech Recognition on TV Series with Video-Guided Post-Correction

Yue Zhang*, Haoyuan Yang*, John Hansen (*Equal Contribution)

Under Review by InterSpeech

- Proposed a Video-Guided Post-Correction framework that refines ASR outputs by leveraging contextual information extracted from video content.
- Evaluate our approach on a multimodal TV series dataset and demonstrate that incorporating video-based context significantly reduces Word Error Rate.
- Responsible for ideas, data collection and annotation, experiment design, and paper writing.

A Unified Hallucination Mitigation Framework for Large Vision-Language Models

Yue Chang, Liqiang Jing, Xiaopeng Zhang, Yue Zhang

TMLR 2024

- Proposed a unified hallucination classification and mitigation framework for Large Vision-Language Models, distinguishing treatments based on classification and utilizing a validation loop for complete hallucination removal.
- Integrated easily into various Large Vision-Language Models, providing flexibility for adding new classifications and treatments.
- Comprehensively evaluated the framework on several benchmarks (MMbench, POPE, CHAIR, and LLaVA-QA90), demonstrating its effectiveness.
- Responsible for idea discussion, framework design, evaluation, and initial paper writing.

Can Large Vision-Language Models Understand Sarcasm?

Xinyu Wang, Liqiang Jing, Yue Zhang

Under Review

- Evaluated Large Vision-Language Models (LVLMs) in multimodal sarcasm analysis tasks, demonstrating their generalization ability across modalities without task-specific fine-tuning.
- Proposed a multi-source semantic-enhanced multimodal sarcasm understanding framework that improves LVLMs' sarcasm understanding by incorporating external knowledge sources.
- Responsible for idea discussion, framework design, and initial paper writing.

Ongoing Projects

Improve Large Video-Language Model for Audio and Video Understanding

Yue Zhang, Liqiang Jing, Vibhav Gogate

- Proposing novel vision model for improving visual representation in Large Vision-Language Models (LVLMs), and improving audio encoder for multilingual scenarios and fundamental audio tasks. Explore reinforcement learning for large multimodal models.

Awards

President's Special Award	2020
Government Scholarship Awarded by the Education Department of Zhejiang Province	2018
First-class Scholarship of Zhejiang Normal University	2017, 2018, 2019

Technical Skills

Languages: Python, Java, C++, R, SQL

Technologies: PyTorch, scikit-learn, Huggingface, PyTorch-Lightning, Weights&Bias, SpaCy, NLTK, Spring Boot, LaTeX

Core Courses: Natural Language Processing, Artificial Intelligence, Machine Learning, Computer Vision